ISBN 978-987-688-054-1

UniRío editora

Introducción a la estadística para las ciencias de la vida

Elsa Moschetti, Susana Ferrero
Gabriela Palacio y Marcelo Ruiz





Introducción a la Estadística para las Ciencias de la Vida

Elsa Moschetti
Susana Ferrero
Gabriela Palacio
Marcelo Ruiz





Universidad Nacional de Río Cuarto
Río Cuarto - Córdoba - Argentina



Uni. Tres primeras letras de “Universidad”. Uso popular muy nuestro; la Uni.
Universidad del latín “universitas” (personas dedicadas al ocio del saber),
se contextualiza para nosotros en nuestro anclaje territorial y en la concepción
de conocimientos y saberes construidos y compartidos socialmente.
El río. Celeste y Naranja. El agua y la arena de nuestro Río Cuarto
en constante confluencia y devenir.
La gota. El acento y el impacto visual: agua en un movimiento
de vuelo libre de un “nosotros”.
Conocimiento que circula y calma la sed.

Introducción a la estadística para las ciencias de la vida /
Elsa Moschetti ... [et.al.]. - 1a ed. - Río Cuarto : UniRío Editora, 2013.
E-Book.

ISBN 978-987-688-054-1

1. Estadística. 2. Biología. 3. Matemática. I. Moschetti, Elsa
CDD 310.4

Fecha de catalogación: 22/10/2013

Introducción a la Estadística para las Ciencias de la Vida
Elsa Moschetti*, *Susana Ferrero*, *Gabriela Palacio y Marcelo Ruiz

Ruta Nacional 36 km 601 – (X5804) Río Cuarto – Argentina
Tel.: 54 (358) 467 6309 – Fax.: 54 (358) 468 0280
editorial@rec.unrc.edu.ar - www.unrc.edu.ar/unrc/comunicacion/editorial/

Primera edición: *Octubre de 2013*

ISBN 978-987-688-054-1



PRÓLOGO

Este libro es el resultado, por un lado de la experiencia didáctica adquirida por los autores al brindar cursos de estadística de grado y posgrado en el área biológica; por otro lado también es el fruto de muchos años de asesoramiento estadístico a investigadores de esta universidad.

El mismo consta de 10 Capítulos, en los cuales se desarrollan los conceptos y técnicas estadísticas de tal forma que cualquier lector con conocimientos elementales de matemática pueda entender la utilidad de las herramientas estadísticas en su área específica. La idea principal se basa en presentar un problema relacionado con el área de la biología e introducir los conceptos estadísticos necesarios para la resolución del mismo. Todos los problemas, ejemplos y ejercicios presentados en este libro han sido seleccionados de manera que resulten de interés para el estudiante y profesionales del área de las ciencias biológicas. Los mismos fueron proporcionados por profesores y/o profesionales del área biológica y otros resultaron de los trabajos de asesoramiento.

El objetivo que se desea alcanzar con este texto es el de proveer al eventual lector la metodología estadística elemental adecuada para la resolución de los diferentes problemas (que comúnmente se les presentan a los investigadores del área biológica) e interpretar la solución de los mismos. Por las razones expuestas y a pesar de que existen muchos "paquetes" estadísticos que permiten resolver rápidamente los cálculos involucrados en la resolución de un problema, en esta primera edición no hemos contemplado la resolución de los ejercicios por medio de ellos.

En este texto, se desea, de alguna manera, poner en relevancia la utilidad de la estadística en situaciones en las que hay que tomar decisiones.

Queremos agradecer a todos los docentes que contribuyeron con material y/o sugerencias para la realización de este libro, así como a todos aquellos que de alguna manera incentivaron nuestro trabajo y confiaron en nuestro esfuerzo. Finalmente, queremos agradecer a la Universidad Nacional de Río Cuarto que hizo posible esta publicación.

Los Autores

INDICE

4 . VARIABLES ALEATORIAS DISCRETAS

5 . VARIABLES ALEATORIAS CONTINUAS

6 . ESTIMACIÓN PUNTUAL Y POR INTERVALOS

7 . PRUEBA DE HIPÓTESIS

8 . INTRODUCCIÓN AL ANÁLISIS DE LA VARIANZA

9 . CORRELACIÓN Y REGRESIÓN LINEAL

10 . PRUEBA DE JI-CUADRADO

APÉNDICE

1 Tratamiento de Datos

Objetivos:

- Reconocer población, muestra, unidad experimental.
- Identificar distintos tipos de variables.
- Resumir e interpretar la información muestral en tablas y gráficos.

1.1 Introducción

La Ciencia, en general, avanza por dos metodologías fundamentales: *deducción* e *inducción.* Con la deducción, a partir de ciertos principios básicos y mediante razonamientos lógicamente correctos, se va tratando de obtener consecuencias y proposiciones que constituyen la teoría. La inducción científica, procede por otro camino bien diferente: partiendo de hechos y observaciones experimentales, trata de llegar a conclusiones generales sobre el objeto que estudia.

Thomas Bayes, en 1763, fue el primero en introducir elementos matemáticos en este proceso inductivo, dando así los pasos iniciales en lo que ha llegado a ser la estadística actual.

Así, la Estadística es una rama del Conocimiento Científico que se ocupa del estudio de las mejores formas de agrupar y analizar datos y de establecer conclusiones acerca del conjunto del que se han recogido tales datos.

Un poco de Historia ...

Desde la antigüedad, reyes y emperadores se preocuparon por conseguir datos sobre sus posesiones. El Imperio Romano, establecido en el año 27 antes de Cristo, fue el primer régimen político que recogió una gran cantidad de datos sobre la población, superficie y bienes de todos los territorios bajo control. Pero hasta comienzos del siglo XVII la estadística era puramente descriptiva, es decir, una enumeración sistemática y ordenada de datos.

Sin embargo, la palabra estadística para designar la obtención, el estudio y la interpretación de grandes masas de datos, parece que fue utilizada por primera vez un siglo más tarde (a mediados del XVIII) en Alemania.

En la segunda mitad del siglo XIX comienza un período de creación y aplicación de técnicas que permiten “inferir” el comportamiento de fenómenos a partir de estudios experimentales.

La estadística se constituyó paulatinamente en una ciencia independiente a principios del siglo XX con los trabajos de los británicos Karl Pearson, sobre los mecanismos de la evolución y herencia, y de R.A. Fisher, con sus estudios sobre tecnología agrícola. Posteriormente la estadística se ha convertido en una base científica esencial para todas las ciencias.

De ningún modo se pretende escribir la historia de la evolución de la estadística sino sólo dar una breve idea del abismal cambio entre sus orígenes y su estado actual.

Algunos de los campos de aplicación de la metodología estadística son: Biología, Agronomía, Veterinaria, Zootecnia, Medicina, Ingeniería, Física, Ciencias Sociales, etc..

La noción de estadística se derivó originalmente del vocablo "estado" porque ha sido función tradicional de los gobiernos centrales llevar registros de población, nacimientos, defunciones, cosechas, impuestos y muchas otras actividades. Contar y medir estos hechos genera muchos datos numéricos.

Las personas no relacionadas con la actividad científica conciben a la estadística como columnas de cifras o gráficos asociados con, por ejemplo, promedios, índices de divorcio, precios de acciones, exportaciones, importaciones.

Este concepto se aproxima mucho a la definición tradicional de estadística la cual es “la compilación, organización, resumen, presentación y análisis de datos numéricos ".

En realidad la función principal de la estadística es elaborar métodos y procedimientos que ayuden a tomar decisiones frente a la incertidumbre, es decir que, además de organizar, analizar y presentar información también la interpreta.

Como procedimiento de toma de decisiones, la estadística se emplea en toda clase de estudios científicos.

Los métodos científicos se utilizan para contestar preguntas tales como: ¿Es efectiva la nueva dieta?, ¿Es eficaz el nuevo medicamento para el dolor de cabeza ?

En realidad, la estadística ha llegado a ser una "herramienta" para todos aquellos profesionales que se ponen en contacto con datos observacionales o experimentales o bien para quienes utilizan los resultados estadísticos determinados por otros. Tales personas necesitan tener alguna familiaridad con principios estadísticos para evitar el mal uso de la misma o la mala interpretación de los resultados generados por ella.

En los últimos años el masivo acceso a las computadoras ha facilitado la implementación y aplicación de métodos estadísticos que permiten describir y ensayar nuevos productos e ideas. Por ejemplo, los médicos estudian los datos obtenidos en los experimentos para desarrollar nuevas medicinas y poner a prueba su eficacia, el gobierno de nuestro país emplea dinero y personas para recolectar y analizar datos a través del Instituto Nacional de Estadísticas y Censos (I.N.D.E.C.).

De muchas maneras se emplea a la estadística para conocer lo que acontece y lo que pueda suceder en el futuro.

> *La Estadística se ocupa entonces de la recolección de datos para descubrir, a través de ellos, nuevos hechos o sea para producir nuevas conclusiones e ideas.*

Generalmente el investigador formula un *problema* de su especialidad y luego junto al estadístico lo transcribe al lenguaje estadístico y, una vez realizado el análisis de los datos, traduce los resultados obtenidos en términos del problema planteado. En lo que se refiere al lenguaje estadístico, existen algunas palabras, como por ejemplo *Población* y *Muestra*, que tienen una acepción muy particular, lo cual requiere que se las especifique con cuidado y que se diferencie su utilización de la dada en el lenguaje cotidiano.

1.2 Formulación de problemas y algunas definiciones fundamentales

Se comenzará formulando problemas, de tipo experimental a partir de los cuales se intentará deducir algunos conceptos útiles.

Formulación de problemas:

1.1 Se propone una nueva dieta para novillos.

1.2 Un laboratorio farmacéutico presenta un nuevo medicamento para aliviar el dolor de cabeza.

1.3 Se trata de imponer una variedad de maíz resistente al Mal de Río Cuarto.

1.4 Una aceitera desea determinar el grado de toxicidad de un lote de semillas de girasol destinadas a la exportación.

En cada situación planteada el problema se reduce a imponer un nuevo "tratamiento", entendiendo por tratamiento a una nueva situación.

A continuación se analizan los distintos problemas:

Problema 1.1: En esta situación se trata de imponer una *nueva dieta*, por lo que surgen naturalmente algunas preguntas como, ¿a quiénes?, ¿a todos los novillos del país?, ¿en toda la provincia?, ¿a novillos de cualquier raza?, ¿hay que tener en cuenta el peso inicial?. En primer lugar se le dará la dieta a novillos del mismo peso inicial y no de cualquier raza, sino a novillos de raza Charolais (raza sobre la que se desea imponer la nueva dieta), por tanto se está restringiendo el efecto de la dieta a un conjunto especial "*novillos de la raza Charolais, con un cierto peso inicial*". A este conjunto se lo denomina *Población de Unidades*.

Problema 1.2: El laboratorio farmacéutico desea imponer en el mercado un nuevo medicamento para disminuir el dolor de cabeza en adultos. Pero ¿este medicamento será efectivo para mujeres?, ¿para varones?, ¿para personas con algún síntoma especial?. Dado que el medicamento es para adultos se debe definir claramente a partir de que edad se considera a una persona adulta, además no se hace distinción de sexo ni de ningún síntoma inicial, por lo tanto se puede pensar que este medicamento está destinado a aliviar el dolor de cabeza en "*adultos (hombres, mujeres) mayores de una cierta edad*". Este conjunto especial se llama *Población de Unidades*.

Problema 1.3: Se desea mejorar en la zona de Río Cuarto el rendimiento de maíz para lo que se buscó un híbrido resistente al Mal de Río Cuarto, el cual afecta el rendimiento de dicho cultivo. Las preguntas ahora son ¿cuál es la zona sobre la que se recomendará este híbrido?, ¿en todo el país?, ¿en la provincia de Córdoba?, ¿en el Departamento Río Cuarto?. Si se decide probarlo en el Departamento Río Cuarto, el conjunto de "*todas las parcelas del Departamento Río Cuarto que podrían ser sembradas con la nueva variedad*" será la *Población de Unidades*.

Problema 1.4: Para determinar el grado de toxicidad de semillas de girasol, almacenadas en silos de una aceitera, se elige al azar un silo de semillas de girasol con ciertas características. Así, las preguntas que se pueden realizar son: ¿el silo tiene almacenadas semillas de la misma zona agrícola?, ¿todos los silos son semejantes, en cuanto a su construcción? ¿todos son conservados de la misma manera? Luego, en este caso se puede considerar a la *Población de Unidades* como el conjunto de "*todos los grupos de semillas de girasol que conforman los silos*" de la aceitera en estudio.

Antes de definir formalmente la población de unidades es necesario dar el concepto de unidad experimental.

Definición 1: La *Unidad Experimental* es el mínimo objeto de estudio sobre el cual se realiza la medición cuantitativa o cualitativa.

En cada uno de los problemas planteados se define la unidad experimental, así se tiene: 1) un novillo de raza Charolais con un determinado peso inicial, 2) un adulto con dolor de cabeza, 3) una parcela del Departamento de Río Cuarto que se podría sembrar con la variedad en estudio, 4) un grupo de semillas extraído con un calador de silos de la aceitera en estudio.

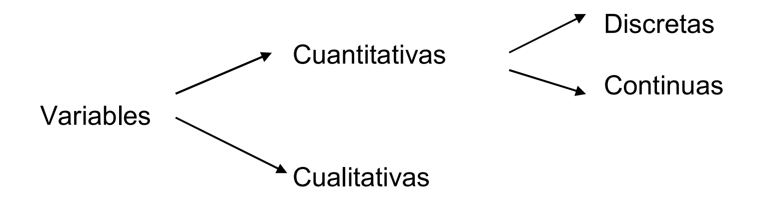
Definición 2: La *Población de Unidades* es el conjunto de unidades experimentales con una característica cualitativa o cuantitativa en común sobre las cuales se extraen las conclusiones del análisis.

En cada uno de los problemas planteados la población de unidades es: 1) todos los novillos de raza Charolais con un determinado peso inicial, 2) todos los adultos con dolor de cabeza, 3) todas las parcelas del Departamento Río Cuarto que se podrían sembrar con maíz, 4) todos los conjuntos posibles de semillas que se pueden obtener con el calador de silos de la aceitera en estudio.

Definición 3: *Variable* es una propiedad objetiva con respecto a la cual las unidades experimentales de la población difieren de manera apreciable. Es la característica que se le "mide u observa" a la unidad experimental, después de haber sido sometida al tratamiento.

En los problemas planteados anteriormente, las variables son: *peso, tiempo, rendimiento y grado de toxicidad*, respectivamente.

En general las variables, de acuerdo a su naturaleza, pueden ser clasificadas en:



1- Cuantitativas (o Medibles): Son aquellas cuyos diferentes estados se pueden expresar con números. Ellas a su vez pueden clasificarse en *Discretas* y *Continuas*.

- **Discretas:** Una variable se considera discreta cuando los valores que asume pasan de un valor a otro consecutivo, sin que pueda tomar valores intermedios.

Ejemplo 1: Número de plantas atacadas, número de hojas, número de gusanos, número de bacterias, número de crías por parición.

- **Continuas:** Una variable se considera continua cuando los valores que asume pueden tomar cualquier valor real comprendido entre dos valores dados.

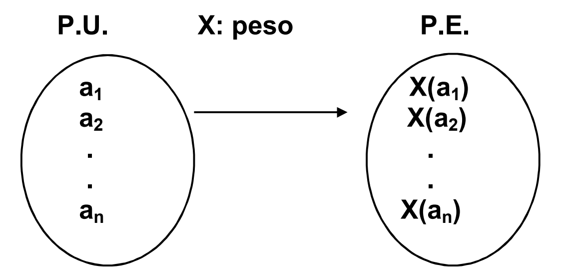
Ejemplo 2: Peso, altura, longitud, ganancia de peso, temperatura, tiempo.

2- Cualitativas: Son aquellas cuyos diferentes estados se expresan por medio de categorías o cualidades.

Ejemplo 3: Color de ojos, color de pelaje, sexo, raza.

Definición 4: Se llama *Población Estadística* al conjunto de todos los valores que resultarían de “medir” la variable, luego de aplicar el tratamiento a las unidades experimentales que forman la población de unidades.

Gráficamente se tiene



Puestos en el lugar del investigador que desea imponer en cada caso un nuevo tratamiento, ¿será conveniente someter a todas las unidades experimentales al tratamiento para determinar si éste es efectivo?.

La respuesta a esta pregunta es si. Pero por otro lado la inspección de todas las unidades experimentales es imposible o poco práctico por razones de tiempo y costo, a menos que la población en estudio sea pequeña.

Pero entonces, ¿la solución está en someter sólo a una unidad experimental al tratamiento en estudio?, éste tampoco es el camino correcto pues con la respuesta de sólo una unidad experimental no se puede tomar una decisión válida. La solución está en tomar un subconjunto representativo de la población en estudio. Para determinar cuántas y cuáles unidades experimentales deberán pertenecer a este subconjunto, se utilizan algunas técnicas estadísticas que no serán abordadas en este texto (Steel, R. y Torrie, J. -1980).

De ahora en más se trabajará con los conceptos sobre uno de los problemas planteados.

Se desea imponer una nueva dieta en novillos de la raza Charolais, con un determinado peso inicial.

Como no se puede probar el tratamiento en sólo una unidad experimental ni tampoco en toda la población se trata de encontrar un *conjunto representativo* donde hacerlo. El problema ahora es cómo *generar* dicho conjunto para que sea representativo de la población en estudio. Una forma de hacerlo para que las técnicas estadísticas puedan ser aplicadas es la siguiente:

Elegir n novillos al azar de la raza Charolais de un cierto peso inicial.

Definición 5: Cada vez que se seleccionan al azar **n** unidades experimentales para luego aplicarles un tratamiento se dice que se realiza un *Experimento Aleatorio*.

En el Problema 1.1 las **n** unidades experimentales son los **n** novillos de la raza Charolais con un cierto peso inicial.

Definición 6: Al conjunto de unidades experimentales seleccionadas se lo denomina *Muestra de Unidades*.

Luego, se puede decir que una *Muestra* es un *subconjunto* de una *Población*.

Una rama de la Estadística estudia acerca de cuántas y cuáles unidades experimentales deben ser seleccionadas para generar una *muestra representativa* de la población en estudio. Cuando en estadística se dice que una muestra es representativa, significa que es un conjunto que reproduce la población en estudio y por tanto puede ser utilizado para conocer alguna característica de la misma.

Definición 7: Al conjunto de resultados obtenidos al "medir u observar" la variable a cada elemento de la muestra de unidades se la denomina *Muestra Estadística*.

Notar que éste es el conjunto de valores de variable utilizado para realizar los análisis estadísticos.

Retomando el Problema 1.1, se tiene

Muestra de Unidades: Los **n** animales de la raza Charolais con cierto peso inicial a los que se les va a suministrar la nueva dieta. Esto se puede expresar como

$$(\text{animal}_1, \text{animal}_2, ..., \text{animal}_n)$$

Variable: Peso. **Tipo:** Cuantitativa continua.

Muestra Estadística: Los **n** valores de peso correspondientes a los animales a los que se les aplicó la dieta. Esto se puede expresar como

$$(\text{Peso}_1, \text{Peso}_2, ..., \text{Peso}_n)$$

Puede ocurrir que más de una unidad experimental tenga el mismo peso, por ello es que se da la siguiente definición.

Definición 8: Se llama *frecuencia absoluta* de un valor de variable X_i al número, f_i, de veces que este valor aparece en la muestra.

Una vez obtenida la muestra estadística, es conveniente resumir la información de la misma. Para ello se utilizan tablas, gráficos y valores descriptivos, a todo lo cual se lo llama ***Estadística Descriptiva***.

1.3 Tablas y Gráficos

1.3.1 Tablas

Ejemplo 4: Suponga que 15 novillos de cierto peso inicial de raza Charolais son sometidos a la nueva dieta y se obtienen los siguientes pesos en kg.

530 498 544 498 532 560 582 560
560 532 544 532 532 582 560

Esta información puede ser resumida en una ***tabla de frecuencias no agrupadas***.

Tabla 1: Número de animales de acuerdo al peso

X_i:Peso	f_i	f_{ri}	F_{ai}
498	2	2/15	2
530	1	1/15	3
532	4	4/15	7
544	2	2/15	9
560	4	4/15	13
582	2	2/15	15
	15	1	

X: variable en estudio, peso.

X_i: diferentes valores de la variable (valores de peso).

f_i: frecuencia absoluta del valor de variable X_i (N$^{ro.}$ de novillos con peso X_i).

f_{ri}: frecuencia relativa del valor de variable X_i (Proporción de novillos con peso X_i).

En general:

k: indica el número de valores distintos de la variable, en este caso $k=6$ y
n: indica el tamaño de la muestra, en este caso es $n=15$ (en general $k \leq n$), donde $n=f_1+f_2+...+f_k$, lo que puede expresarse como $n = \sum_{i=1}^{k} f_i$.

Definición 9: Se llama *frecuencia relativa* (f_{ri}) del i-ésimo valor de variable X_i a la frecuencia absoluta dividida por el tamaño de la muestra, en símbolos

$$f_{ri} = \frac{f_i}{n}$$

Definición 10: La *frecuencia acumulada* (F_{ai}) correspondiente al valor X_i es la suma de las frecuencias absolutas de los valores de variable *menores o iguales* a X_i.

La información de la Tabla 1 se puede leer, por ejemplo, de la siguiente manera:

- la frecuencia absoluta indica que 4 novillos tuvieron un peso de 532 kg.
- la frecuencia relativa indica que de los 15 animales 4 tuvieron un peso de 532 kg., o que aproximadamente el 27% de los animales pesan 532 kg.
- la frecuencia acumulada dice que 7 animales alcanzaron 532 kg. o menos.

Si hubiera muchos valores diferentes de variable, esta tabla no sería adecuada para resumir la información.

Ejemplo 5: Suponga que 20 novillos de la raza Charolais con un determinado peso inicial son sometidos a la nueva dieta. Los resultados obtenidos son:

490	498	499	500	532	531	518	516	540	561
555	566	562	603	602	610	612	612	525	583

Si se trata de construir una tabla como la anterior, se podrá observar que la misma no resume la información de la muestra.

Es por ello que surge la necesidad de construir otro tipo de tablas, en las cuales se agrupan los valores de variable en intervalos.

Así en este caso se procede como se indica a continuación:

Dado que se deben construir intervalos, hay que tener en cuenta la cantidad y la longitud conveniente de cada uno. En base al Ejemplo 5, una regla práctica para construirlos es:

1. Usar $k=5$, donde **k** indica el número de intervalos.

2. $X_{máx}=612$ y $X_{mín}=490$

3. $h = \frac{(X_{max} - X_{min})}{5} = \frac{612 - 490}{5} = 24.4 \cong 25$, donde $\mathbf{h}$ indica la longitud de cada intervalo. El resultado de $\mathbf{h}$ se redondea siempre <u>por exceso</u> y debe tener la misma cantidad de decimales que los datos.

4. Ahora, como $X_{mín}+h = 490+25 = 515$ entonces el primer intervalo es $[490,515)$, que incluye al valor 490 y no al valor 515; el segundo intervalo y los siguientes quedan como se observa en la tabla. El último es siempre un intervalo cerrado.

A continuación se muestra la ***tabla de frecuencias agrupadas*** para los datos del Ejemplo 5.

Tabla 2: Número de animales de acuerdo al peso

Intervalo de Clase	Conteo	f_i	f_{ri}
[490, 515)	||||	4	4/20
[515, 540)	|||||	5	5/20
[540, 565)	||||	4	4/20
[565, 590)	||	2	2/20
[590, 615]	|||||	5	4/20
		20	1

Las frecuencias f_i y f_{ri} representan, respectivamente, a la frecuencia absoluta y relativa del i-ésimo intervalo.

Para completar la descripción de estos datos se debe realizar otra tabla con las frecuencias acumuladas, que será descripta en la Sección 1.3.2 (Tabla 4).

Para construir estos intervalos se han tenido en cuenta algunas características, tales como:

1. El primer intervalo de clase debe contener al valor mínimo y el último al máximo.

2. La cantidad de intervalos debe aumentar a medida que aumenta $\mathbf{n}$. El número $\mathbf{k}$ de intervalos aconsejable de acuerdo al tamaño de la muestra es:

$$\begin{array}{rcl} n \leq 50 & \rightarrow & k=5,6 \\ 50 < n \leq 100 & \rightarrow & k=6,7 \\ 100 < n \leq 500 & \rightarrow & k=7,8 \text{ o } 9 \\ 500 < n \leq 2000 & \rightarrow & k=10,11,12 \\ n > 2000 & \rightarrow & k=13,14,...,20 \end{array}$$

Cuando se trabaja con este tipo de tabla se toma como representante de cada intervalo al punto medio del mismo, el que recibe el nombre de *Marca de clase* del intervalo, se denota por $\widetilde{X}_i$ para el intervalo $\mathbf{i}$ y se calcula como

$$\widetilde{X}_i = \frac{X_{Lím.Inferior} + X_{Lím.Superior}}{2}$$

Notar que la elección de la tabla para resumir la información de la muestra estadística no depende solamente de la variable en estudio, sino también del tamaño de muestra y de las frecuencias, pues si muchos valores de variable son coincidentes una tabla de frecuencias no agrupadas seguramente resume muy bien la información de la muestra, en tanto que si los valores de variable no se repiten, entonces una tabla de frecuencias agrupadas es la adecuada.

Definición 11: La *frecuencia acumulada* (F_{ai}) correspondiente al i-ésimo intervalo es la suma de las frecuencias absolutas de ese intervalo con las frecuencias de los anteriores.

La información de la Tabla 2 se puede leer, por ejemplo, de la siguiente forma:

- la frecuencia absoluta está indicando que hay 4 animales que entre 540 y 565 kg.
- la frecuencia relativa indica que de los 20 animales 4 tuvieron un peso entre 540 y 565 kg., o que aproximadamente el 20% de los animales pesan entre 540 y 565 kg.

En los dos ejemplos anteriores se han presentado variables cuantitativas, ahora se dará un ejemplo donde la variable observada es de tipo cualitativa.

Ejemplo 6: En una cabaña se desea clasificar a los equinos de una cierta raza y edad de acuerdo al color del pelaje. Para ello se seleccionaron aleatoriamente 20 animales a los que se les observó el color del pelaje clasificándolos en a: alazán z: zaino y t: tordillo, obteniéndose los siguientes datos:

z a t z t t z a a z a t a t a a a z a t

La información de una muestra estadística donde la variable observada es de tipo cualitativo se puede resumir *sólo* en una *tabla de frecuencias no agrupadas.* Luego para este caso se tiene

Tabla 3: Clasificación de equinos según el color del pelaje

X_i (color de pelaje)	f_i	f_{ri}
z	5	5/20
a	9	9/20
t	6	6/20
	20	1

La información de la Tabla 3 se interpreta, por ejemplo, de la siguiente forma:

- la frecuencia absoluta está indicando que hay 5 animales de color zaino.
- la frecuencia relativa esta indicando que de los 20 animales 5 son de color zaino, o que el 25% de los animales de la muestra son de color zaino.

Cuando la variable en estudio es de tipo cualitativo, la frecuencia acumulada no tiene sentido, pues los valores de variables no tienen un orden natural.

1.3.2 Gráficos

Otra forma de presentar la información muestral es a través de gráficos. Éstos permiten, cuando han sido correctamente realizados, obtener en forma rápida una primera idea del comportamiento de los datos a ser analizados.

En la construcción de los gráficos no hay reglas estrictas y generales que deban ser seguidas pero sí es necesario tener en cuenta algunas recomendaciones, como por ejemplo:

a) Aquel gráfico que alcance su objetivo con la máxima sencillez será el más efectivo.

b) La representación gráfica debe ser clara y simple para que con una "mirada" se tenga una idea de la distribución de los datos.

c) Toda representación gráfica debe explicarse por sí misma, para lo cual, deben estar indicados el título, origen y escala (el título debe expresar con claridad y en forma breve aquello que se propone mostrar con el gráfico).

d) Por lo general en el eje de las ordenadas (vertical) se representa la frecuencia y los valores de variable en el eje de las abscisas (horizontal). Las dos escalas que se emplean deben marcarse con toda claridad como así también las unidades en que están expresadas.

e) El eje horizontal no necesariamente debe comenzar en cero pero sí el eje vertical, para evitar estimaciones visuales erróneas.

f) En cuanto al *tamaño* del gráfico se deben elegir escalas adecuadas a la magnitudes que se quieren representar y al tipo de fenómeno que está en estudio, por cuanto el uso de escalas incorrectas puede llevar a falsas ideas acerca del comportamiento de las variables; por ejemplo una escala muy estrecha en las abscisas y muy amplia en las ordenadas puede resultar en un gráfico que magnifique las fluctuaciones de la variable de interés; por el contrario una escala muy amplia en las abscisas y estrecha en las ordenadas dará como resultado un gráfico achatado y suavizará en demasía las fluctuaciones de la variable en estudio.

g) El gráfico debe tener una referencia acerca de la *fuente* de donde provienen los datos que se representan, indicando autor, título, volumen, página, editor y fecha. La misma se ubica en la parte inferior del gráfico.

Los gráficos que serán utilizados aquí para representar la información de las tablas construidas anteriormente son:

a) Para Tablas de Frecuencias no Agrupadas

1. Diagrama de Barras

En un sistema de coordenadas cartesianas se representan en el eje de abscisas los valores de la variable X y en el eje de ordenadas las frecuencias absolutas f_i (o equivalentemente las frecuencias relativas f_{ri}). Sobre cada valor de la variable se levanta una línea o una barra de ancho fijo y altura f_i. Por ejemplo para los datos del Ejemplo 4, el diagrama de barras que corresponde es el siguiente:

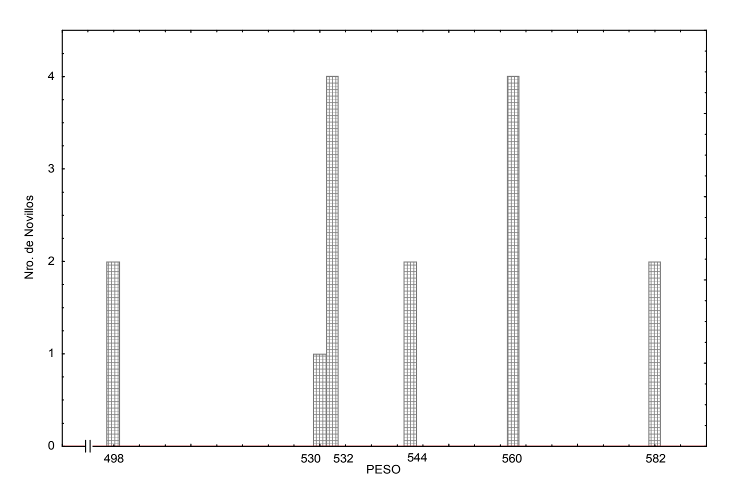


Gráfico 1: Distribución de novillos de acuerdo al peso

2. Polígono de Frecuencias Acumuladas

El polígono de frecuencias acumuladas es una función escalonada que a cada valor de X_i le asigna la frecuencia acumulada F_{ai}. Sobre el eje de abscisas se marcan los valores de

variable y sobre el eje de ordenadas las frecuencias acumuladas. El gráfico para los datos del Ejemplo 4 se muestra a continuación.

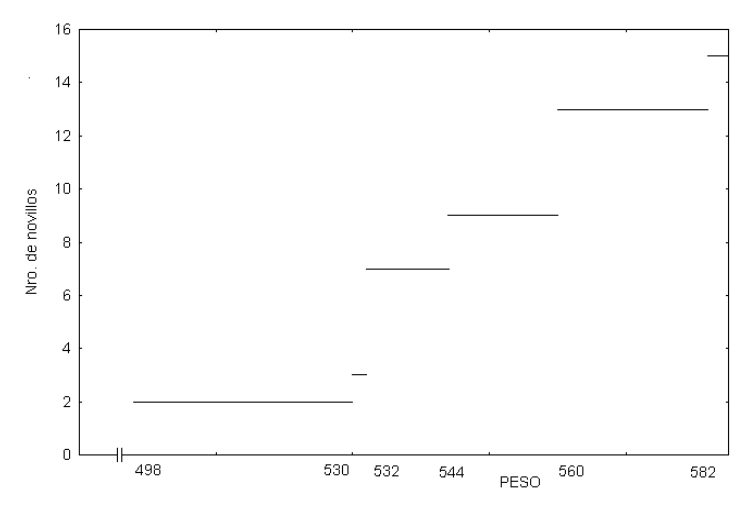


Gráfico 2: Distribución de los novillos de acuerdo al peso

b) Para Tablas de Frecuencias Agrupadas

1. Histograma

En este gráfico se representan los intervalos de clase sobre el eje de abscisas y sobre cada uno de ellos se levanta un rectángulo con ancho fijo y altura igual a la frecuencia absoluta f_i de ese intervalo (o equivalentemente f_{ri}). A continuación se muestra el histograma correspondiente a los datos del Ejemplo 5. Observar que este gráfico se construye en base a la Tabla 2.

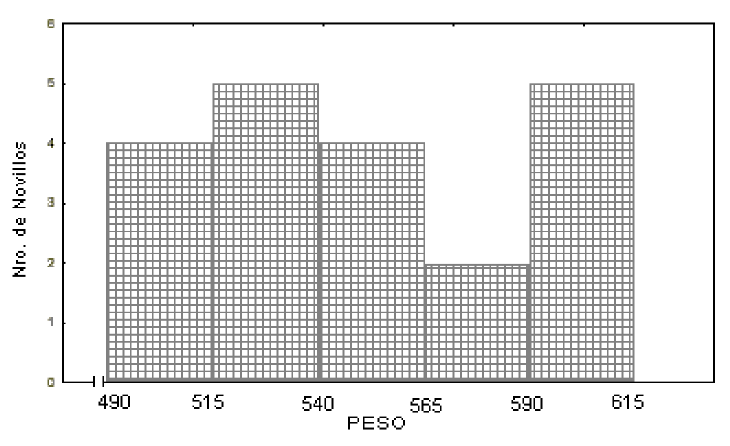


Gráfico 3: Distribución de los novillos de acuerdo al peso

Este gráfico no se puede utilizar cuando los intervalos son de longitudes diferentes; en estos casos se debe usar un Histograma de Áreas, cuyos detalles no serán presentados aquí.

2. Polígono de Frecuencias Acumuladas

Para realizar este gráfico previamente se debe construir la tabla de frecuencias acumuladas (F_{ai}). La misma resulta asignando al límite inferior del primer intervalo el valor cero, al límite superior del primer intervalo el valor f_1, al límite superior del segundo intervalo el valor f_1+f_2 y así sucesivamente, es decir que al límite superior del i-ésimo intervalo se le asigna

el valor $f_1+f_2+...+f_i$. En la Tabla 4 se presentan las frecuencias acumuladas para los datos del Ejemplo 5.

El polígono de frecuencias acumuladas se obtiene uniendo los puntos (t_i, F_{ai}) mediante segmentos de recta, donde t_1 denota el límite inferior del primer intervalo y t_i denota el límite superior del i-ésimo intervalo con $i=2,3,...,k$.

Tabla 4: Número de animales según el peso.

$X<X_i$	F_{ai}
490	0
515	4
540	9
565	13
590	15
615	20

Esta tabla indica que, por ejemplo, no hay ningún animal que pese menos de 490 kg., que hay 13 animales que pesan menos de 565 kg..

El siguiente gráfico representa las frecuencias acumuladas para una tabla de frecuencias agrupadas.

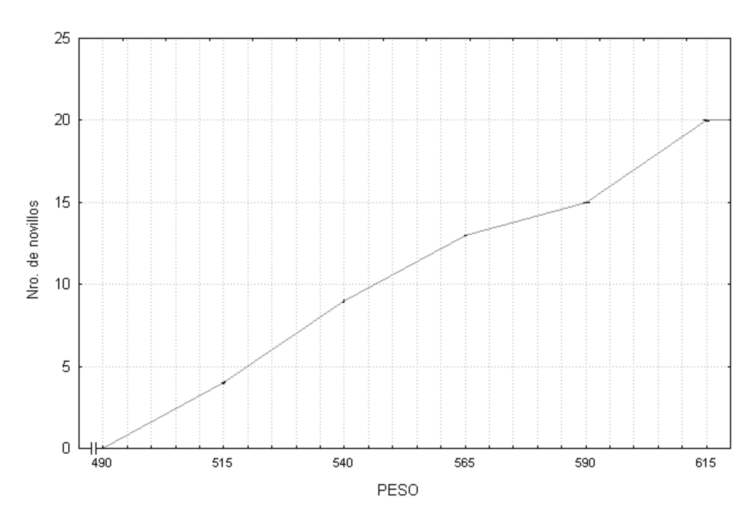


Gráfico 4: Distribución de los novillos de acuerdo al peso

La diferencia entre los dos polígonos de frecuencias acumuladas (Gráficos 2 y 4) es que, cuando la tabla de frecuencias es no agrupada se considera como si entre dos valores consecutivos de la variable no hubiera ningún valor posible, lo que da lugar a un gráfico en forma de escalera; mientras que cuando los datos son resumidos en una tabla de frecuencias agrupadas se asume que la variable puede tomar cualquier valor entre los extremos de cada intervalo, por lo cual el gráfico correspondiente resulta una poligonal.

Notar que cuando la variable en estudio es de tipo cualitativo sólo tiene sentido el diagrama de barras.

c) Diagrama de Tallo y Hojas

Un procedimiento semi-gráfico de presentar la información para variables cuantitativas, que es especialmente útil cuando el número total de datos es pequeño (menor que 50), es el diagrama de tallo y hojas de Tukey. Los principios para construirlo son:

a) Redondear los datos a dos o tres cifras significativas, expresándolos en unidades convenientes.
b) Construir una tabla con dos columnas (separadas por una línea) como sigue:
   b_1) Para datos con dos dígitos, escribir a la izquierda de la línea los dígitos de las decenas, que forman el tallo, y a la derecha las unidades, que serán hojas.
   b_2) Para los datos con tres dígitos el tallo estará formado por los dígitos de las centenas y decenas, que se escribirán a la izquierda y a la derecha de la línea van las unidades.
c) Cada *tallo* define una clase y se escribe sólo una vez. El número de *hojas* representa la *frecuencia* de dicha clase.

Ejemplo 7: Los datos que se presentan a continuación corresponden a la altura en cm. de cierto arbusto.

11.357 12.542 11.384 12.431 14.212 15.213 13.300 11.300 17.206 12.710
13.455 16.143 12.162 12.721 13.420 14.698 11.312 11.217 11.414 11.142

Se realiza un cambio de escala (de cm. a mm.) y luego se redondea a tres cifras significativas, lo cual genera la siguiente muestra estadística:

114	125	114	124	142	152	133	113	172	127
135	161	122	127	134	147	113	112	114	111

El Diagrama de Tallo y Hojas correspondiente es

11	1233444
12	24577
13	345
14	27
15	2
16	1
17	2

Gráfico 5: Distribución de los arbustos según su altura.

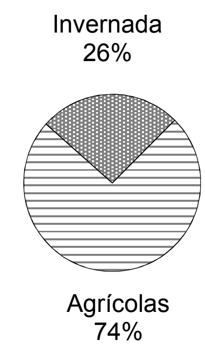
Si bien hay diferentes formas de construir un diagrama de tallo y hojas, el aquí presentado es el más simple.

d) Gráfico Circular o Diagrama de Torta

Se construye mediante la división proporcional de un círculo, de cualquier radio, en sectores circulares para cada clase de acuerdo al porcentaje que cada magnitud representa del total.

EL VALOR DE LA TIERRA A TRAVÉS DE LOS AÑOS EN LA ARGENTINA

Invernada
26%



Agrícolas
74%

Fuente: Bullrich Campos S.A. (extraído del diario La Nación - 20/6/96)

Gráfico 6: Valor de la tierra de acuerdo a su uso en el año 1996.

1.4 Análisis Descriptivo Multivariado

Hasta ahora se han estudiado algunas herramientas de la estadística descriptiva para cuando se desea analizar sólo una variable, lo que es denominado estudio univariado. Cuando se desea estudiar el comportamiento de más de una variable simultáneamente, el análisis estadístico recibe el nombre de análisis multivariado. Lo más común es que se estudien conjuntamente dos variables lo que es llamado un análisis bivariado.

1.4.1 Análisis bivariado para variables cuantitativas

Diagrama de dispersión

Para observar si dos variables de tipo cuantitativas (principalmente continuas) pueden estar relacionadas se realiza un *Diagrama de Dispersión*. Para ejemplificar este tipo de gráfico será presentado la siguiente situación.

Problema 1.5: Suponga que un grupo de investigadores sospecha que hay asociación entre el peso y el volumen sanguíneo de cabras de una cierta raza. Para confirmar su sospecha tomó aleatoriamente 12 cabras de la raza en estudio, de cierto peso y edad (unidad experimental), midiéndole a cada una de ellas las dos variables simultáneamente. Los datos obtenidos se muestran a continuación:

X: Peso (kg.)	34	28	19	41	21	20	21	39	37	23	17	48
Y: Volumen (cm^3)	2.3	2.1	1.1	2.8	1.5	1.6	1.4	2.4	2.5	1.5	1.1	3.5

El Diagrama de Dispersión muestra la ubicación de los pares de observaciones (X,Y) en un sistema de coordenadas cartesianas, como se puede observar en el gráfico siguiente.

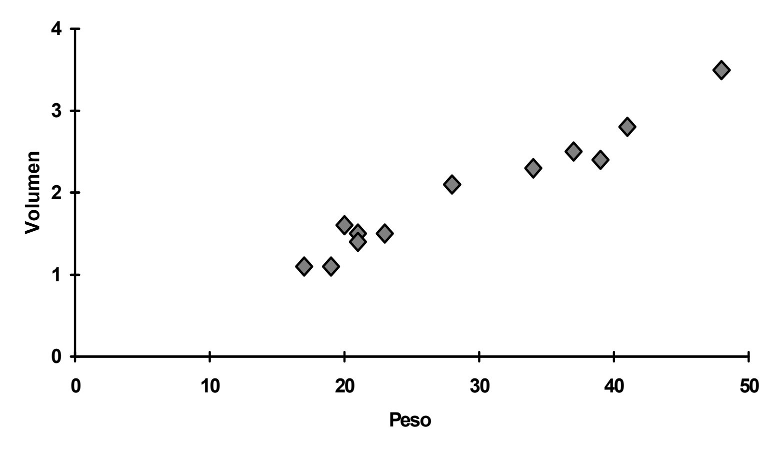


Gráfico 7: Peso y volumen sanguíneo de cabras.

Dado que los puntos se aproximan a una recta de pendiente distinta de cero, el gráfico sugiere que las variables peso y volumen sanguíneo están relacionadas *linealmente*. Como la recta tiene pendiente positiva (a mayor peso corresponde mayor volumen sanguíneo) se puede pensar que existe una *asociación lineal positiva* entre el peso y el volumen sanguíneo.

En general, dos variables de tipo cuantitativo pueden tener:

Asociación Lineal Positiva: si a medida que aumentan los valores de la variable X también aumentan los valores de la variable Y (Gráfico 8-a).

Asociación Lineal Negativa: si a medida que aumentan los valores de la variable X disminuyen los valores de la variable Y (Gráfico 8-b).

Ausencia de Asociación Lineal: cuando los puntos se dispersan en el plano o están distribuidos alrededor de alguna otra curva. (Gráfico 8-c y d).

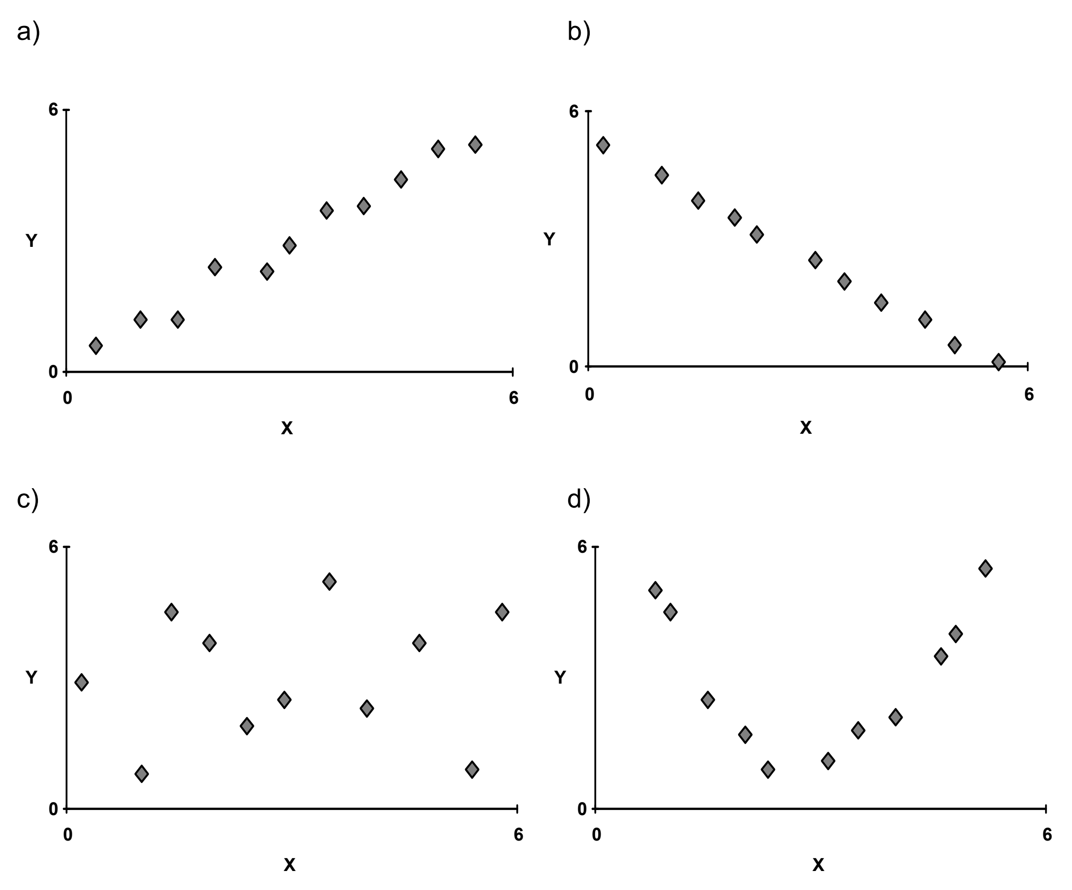


Gráfico 8: Diagramas de dispersión. a) Asociación Lineal Positiva;
b) Asociación Lineal Negativa;
c) y d) Ausencia de Asociación Lineal.

1.4.2 Análisis bivariado para variables cualitativas

Tablas de contingencia o de doble entrada.

Para estudiar problemas de asociación entre dos variables aleatorias cualitativas y para resumir la información, se construyen en primer lugar las llamadas *tablas de contingencia o de doble entrada*. Para ejemplificar este tipo de tablas será presentada la siguiente situación.

Problema 1.6: Un investigador interesado en estudiar tumores cutáneos en equinos, tomó una muestra aleatoria de 1000 equinos y observó la presencia de dichos tumores y el color del pelaje. En la Tabla 5 se resume la información obtenida.

Una tabla de contingencia se construye con las frecuencias correspondientes al combinar las categorías de las variables en estudio.

Tabla 5: Distribución de frecuencias según el color y presencia del tumor.

PRESENCIA DE TUMOR COLOR	Si	No	TOTAL
Alazán	220	80	300
Zaino	135	115	250
Tordillo	415	35	450
TOTAL	770	230	1000

La frecuencia de animales zainos con tumor es $f_{ZS} = 135$[1]. Ésta indica que de los 1000 estudiados 135 son de color zaino y tienen tumores cutáneos.

El gráfico más adecuado asociado a esta tabla es el Diagrama de barras múltiples que se presenta en la sección siguiente.

1.4.3 Otros gráficos

1. Diagrama de barras múltiples

Este gráfico se construye mediante la ubicación de dos o más rectángulos (barras), para cada valor de la variable representado en el eje de abscisas. La altura de cada barra varía según sea la magnitud a representar en el eje de ordenadas que no es necesariamente la frecuencia (como en el diagrama de barras e histograma presentados). En el caso de una tabla de contingencia la magnitud del eje de ordenadas es la frecuencia. Notar que en el eje x se puede representar más de una variable.

Algunas veces las barras se ubican en forma horizontal, cambiando adecuadamente lo que se representa en cada eje.

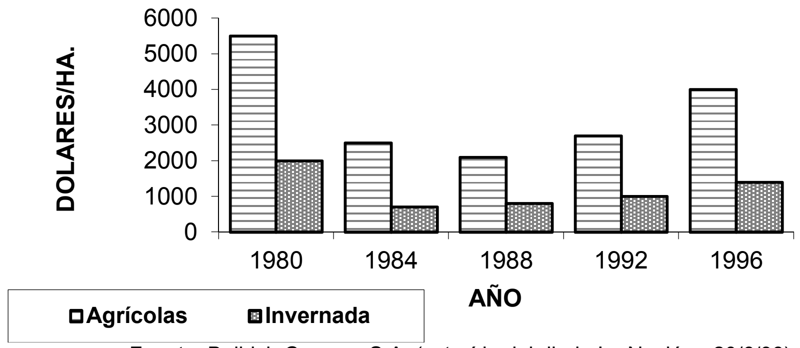


Gráfico 9: Valor de la tierra de acuerdo a su uso a través de los años

Para la Tabla 5 el Diagrama de barras múltiples es

[1] *Esta frecuencia puede denotarse f_{21}, donde el subíndice **2** indica la fila (Zaino) y el subíndice **1** indica la columna (Si).*

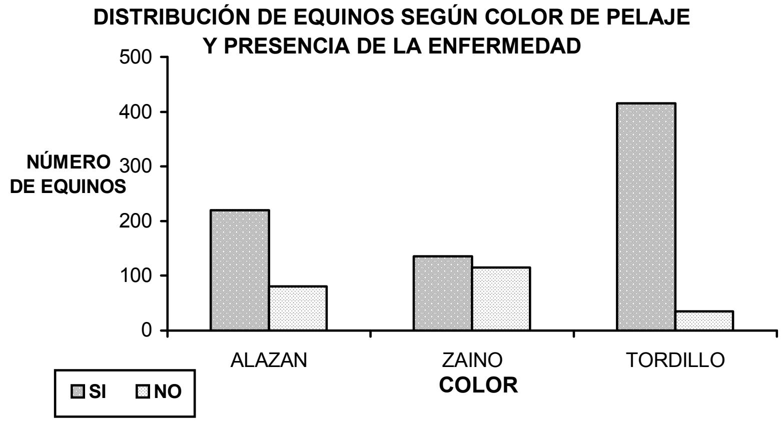


Gráfico 10: Distribución de frecuencias según el color y presencia del tumor.

2. Diagrama de barras componentes

Este gráfico está formado por barras superpuestas en vez de contiguas para cada valor de variable. Como en el tipo de gráfico anterior, la magnitud del eje de ordenadas no es necesariamente una frecuencia, y cuando se trata de una tabla de contingencia sí lo es.

Como en el caso anterior, algunas veces las barras se ubican en forma horizontal, cambiando adecuadamente lo que se representa en cada eje.

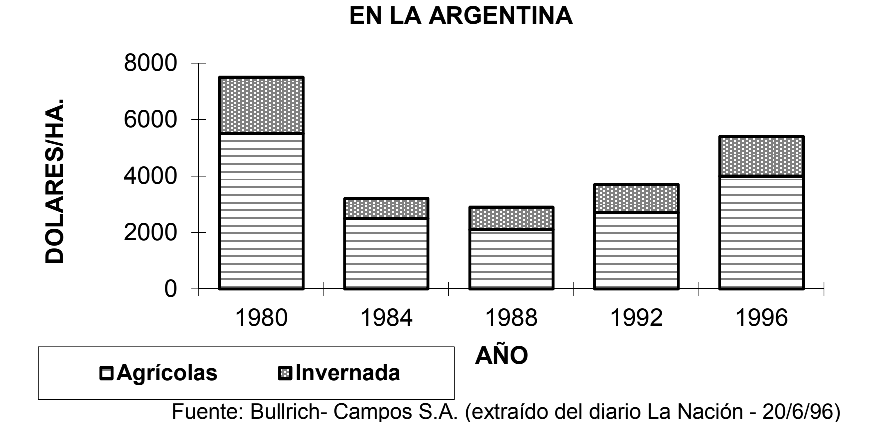


Gráfico 11: Valor de la tierra de acuerdo a su uso a través de los años

3. Diagrama de Líneas Múltiples

Consiste en reemplazar las barras por puntos que se unen con una línea, como puede observarse en el Gráfico 12. Como antes, la magnitud del eje de ordenadas no es necesariamente una frecuencia.

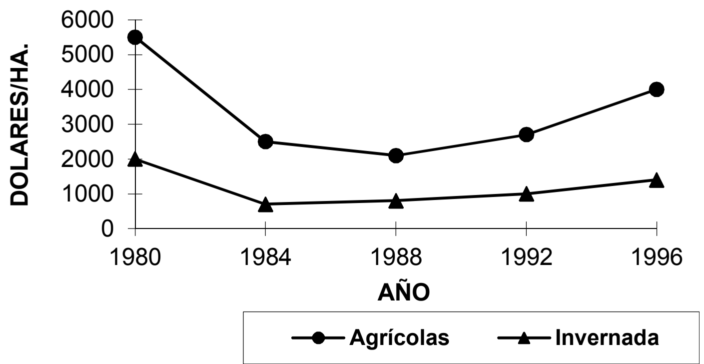


Gráfico 12: Valor de la tierra de acuerdo a su uso a través de los años.

Resolver e interpretar los resultados, de los ejercicios de aplicación propuestos a continuación y al final de cada uno de los capítulos siguientes, permitirá apreciar cuanto pueden ayudar las técnicas estadísticas en la resolución de un problema planteado desde una área aplicada.

Ejercicios de Aplicación

1.

Para cada uno de los siguientes ejemplos indicar la Unidad Experimental, la Muestra de Unidades y la Población de Unidades.

a) Se desea determinar el rendimiento de una nueva variedad de trigo en una zona del sur de la Provincia de Santa Fe. Para ello se seleccionaron 30 parcelas a las que se les midió el rendimiento en tn/ha.

b) Para determinar el peso de los huevos de ponedoras Leghorn blanca de una línea se consideraron 16 huevos de tales aves y se midió el peso de los mismos.

c) Los lechones suelen ser alimentados en recintos separados de los que comen las cerdas para evitar que se molesten. Para estudiar el aumento de peso de lechones alimentados de esa manera, se seleccionaron 20 de ellos y se les midió dicha variable.

d) Para realizar un estudio acerca de la presencia de ciertos parásitos se realizó un análisis coproparasitológico a 36 niños entre 0 y 13 años de edad, aplicándoles cierta técnica de detección de parásitos y registrando la presencia o ausencia de los mismos.

e) A fin de estimar el número de lechones por año en cerdas Yorkshire Holandés se escogió una muestra de 20 cerdas de dicha raza y se contó el número de lechones por cerda.

f) Se quiere estimar la composición botánica en la dieta anual de la vicuña en libre pastoreo en la Puna Jujeña. Se trabajó con 25 animales, recolectando de cada uno 1 muestra de 100 gramos de sus heces. En cada una de las muestras de heces se midió el porcentaje de Festuca Scirpifolia y de Deyeuxia Nardifolia (el resto de la composición de la dieta se clasificó en Otros).

g) Se quiere saber si cierto componente en la dieta de pollos parrilleros disminuye la conversión alimenticia, la cual se obtiene haciendo el cociente entre el alimento consumido y la ganancia de peso, medidas ambas por corral. Para ello se tomaron 50 pollos en el 2º día de vida, se dividieron en corrales de 5 aves cada uno, y al cabo de 1.5 meses de suministrarle la dieta con el nuevo componente se efectuaron las mediciones.

2.

En las situaciones anteriores determinar las variables en estudio, el tipo al cual corresponden, la Muestra Estadística y la Población Estadística.

3.

Los caballos de la raza “sangre pura de carrera” de la Argentina fueron clasificados según su lugar de procedencia (provincia), arrojando los resultados que se muestran en la tabla:

a) ¿Cuál es la unidad experimental en este estudio?.

b) ¿Los datos anteriores corresponden a una población o a una muestra de unidades?.

c) ¿Cuál es la variable estudiada?.

d) Graficar la información presentada en la tabla.

Provincia	Número de Animales
Santa Fe	130
Buenos Aires	800
Corrientes	8
Chubut	8
Santa Cruz	1
Entre Ríos	102
La Pampa	50
Mendoza	9
San Luis	13
Santiago del Estero	11
Río Negro	5
San Juan	8
Tucumán	3
Córdoba	170
Chaco	4

4.

Suponga que los datos correspondientes al inciso e) del Ejercicio 1 son los siguientes:

Número de lechones por año	Número de cerdas
10	3
12	4
14	7
15	4
16	2

a) Completar la tabla con los tipos de frecuencias faltantes.
b) Realizar los gráficos correspondientes.
c) ¿Cuál es el significado de la segunda frecuencia absoluta (f_2)?.

5.

I) Sobre la composición botánica de la dieta de la vicuña, se obtuvieron los siguientes datos respecto de la Festuca (Festuca Scirpifolia) en % (inciso f) del Ejercicio 1:

35	43	52	43	54
78	65	64	76	62
63	53	53.5	57	55.7
61.2	70.3	68.6	68	51
50.5	50.5	53	66.5	72

Resumir este conjunto de valores a través de tablas y gráficos.

II) Los datos referidos a la Deyeuxia (Deyeuxia Nardifolia) se presentan en la siguiente tabla:

Límite del Intervalo	Frecuencia Acumulada
20	0
27	5
34	9
41	14
48	20
55	25

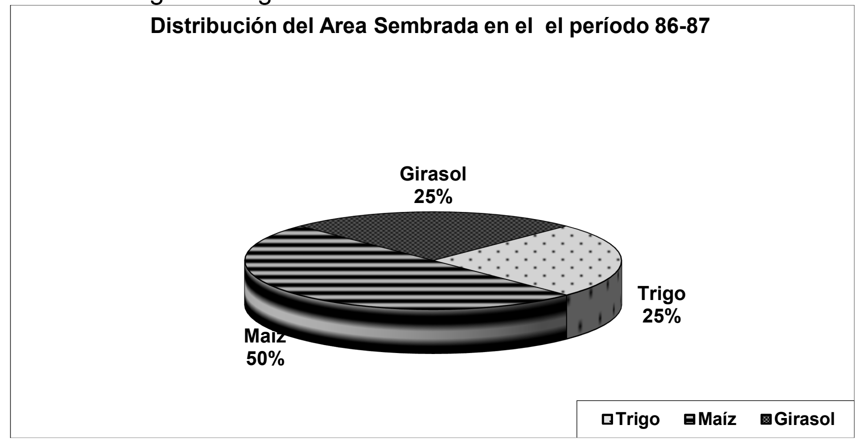
a) ¿Qué cantidad de muestras de heces presentaron un porcentaje de Deyeuxia entre 27% y 34% ?.

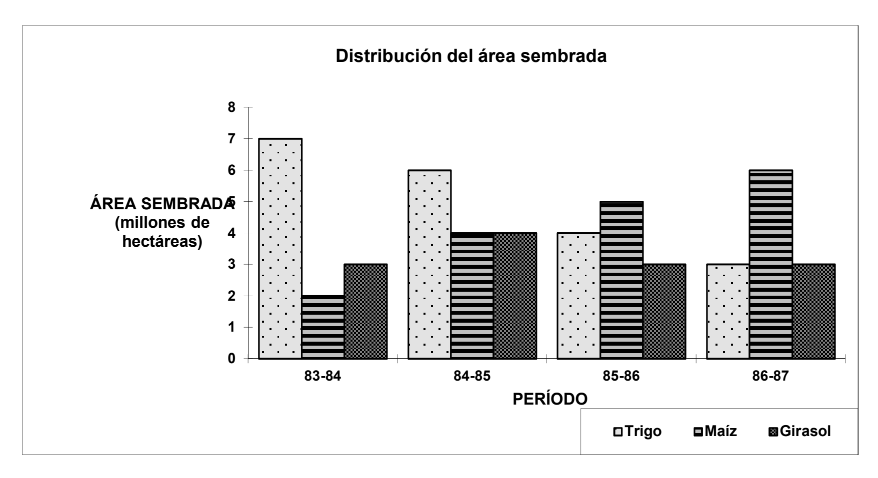
b) ¿Qué cantidad de muestras de heces presentaron un porcentaje de Deyeuxia inferior al 41 %?.

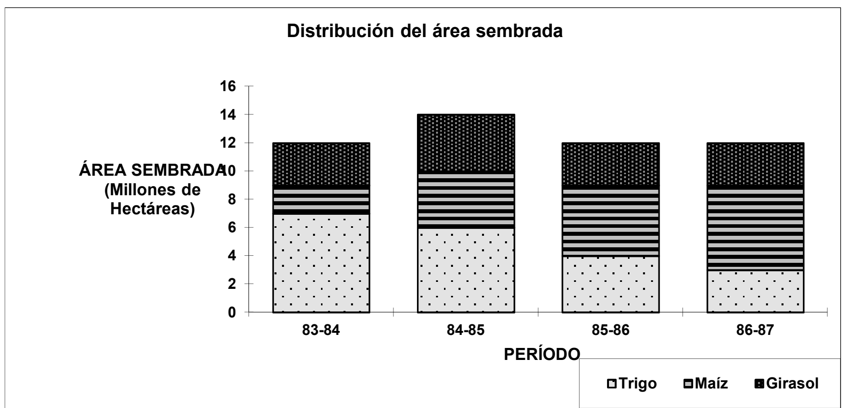
c) Realizar, a partir de la tabla anterior, el histograma y el polígono de frecuencias acumuladas.

6.

En base a los siguientes gráficos:







a) ¿ Cuál es la tendencia del área sembrada de girasol y qué diferencia tiene con la de los otros cultivos?.

b) ¿Cómo son las proporciones de áreas sembradas en relación al área total, en el último período considerado?.

c) Si tuviera que comparar las proporciones de las áreas sembradas de los cultivos en el período 83-84 ¿Qué gráfico utilizaría ?.

d) ¿Qué gráfico elegiría si tuviera que comparar proporciones de cada uno de los cultivos en distintos períodos ?.

7.

La furazolidona, quimioterápico usado en explotaciones aviares, produce efectos tóxicos en numerosos órganos. También se ha observado descenso en la ganancia de peso y consumo de alimento. En un trabajo de un grupo de investigadores de la Universidad Nacional de Río Cuarto se ha estudiado el efecto de la furazolidona sobre la ganancia de peso en pavos híbridos. Para ello se tomó una muestra de pavos machos de 5 semanas de vida. A partir de este momento las aves fueron divididas en dos grupos: durante 12 semanas a un grupo (tratado) se le dio un alimento comercial con agregado de furazolidona (al 0,04%) y al otro (control) el alimento sin el agregado. Al cabo de la 12º semana se efectuaron las mediciones.

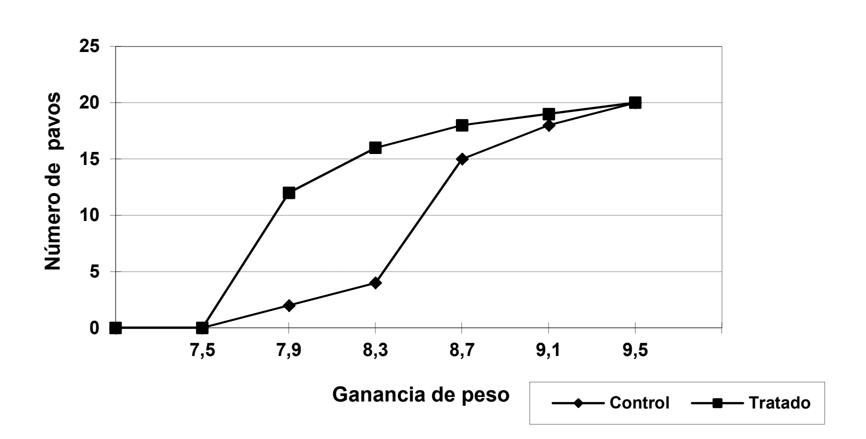
a) En esta experiencia: ¿Cuál es la unidad experimental, la muestra de unidades y la población de unidades?. ¿Cuál la variable, la muestra estadística y la población estadística?.

A partir de la información dada en el gráfico:

I. ¿Cuáles son las mínimas y máximas ganancias de peso consideradas?

II. ¿Qué ocurre cuando en el gráfico se presentan pendientes muy pronunciadas o muy bajas?.
III. ¿Qué cantidad de pavos tratados y controles tienen una ganancia de peso inferior a los 8.3 kg.?. ¿Cuántos tratados y controles tienen una ganancia de peso superior a 8.3 kg.?.

b) De la comparación de ambos polígonos, ¿surge algún indicio acerca de si la furazolidona influye o no sobre la ganancia de peso ?



2 Estadísticos

Objetivos:

- Sintetizar la información de la muestra a través de valores representativos.
- Reconocer la insuficiencia de los estadísticos de posición como únicas medidas descriptivas de una muestra.
- Seleccionar los estadísticos que mejor describen una muestra.
- Interpretar la información brindada por los estadísticos para una situación particular.

2.1 Introducción

Hasta ahora se ha logrado resumir la información muestral a través de tablas y gráficos. A continuación se verá otra manera de caracterizar una muestra, presentando algunos ejemplos.

Ejemplo 1: Retomando la situación presentada en el Problema 1.1, suponga que a 7 animales de un determinado peso inicial de la raza Charolais elegidos al azar se les aplica la nueva dieta. El peso obtenido por cada uno de ellos es

muestra$_0$: 200 300 300 400 500 500 600

¿Qué número se puede calcular para tener una idea general del efecto de la nueva dieta?, es decir ¿con qué valor numérico se podrá caracterizar esta muestra de peso de animales sometidos a la nueva dieta?. Parece natural tomar el *promedio*, el cual se calcula de la siguiente manera

$$\overline{X} : \text{promedio} = \frac{200+300+300+400+500+500+600}{7} = 400$$

Ejemplo 2: Se toman tres muestras de 7 animales cada una de la raza Charolais y se les aplica la nueva dieta, obteniéndose los siguientes pesos.

muestra$_1$: 400 450 500 550 600 680 700 $\overline{X}$ = 554.29 kg.

muestra$_2$: 200 200 200 200 240 270 700 $\overline{X}$ = 287.14 kg.

muestra$_3$: 180 190 210 230 250 280 700 $\overline{X}$ = 291.43 kg.

Observar si el promedio (también llamado media) es realmente representativo en cada una de las muestras, es decir si da una idea real de lo que sucede:

En la primera de ellas sí, pero en las otras dos no. En la **muestra**$_1$ hay 4 animales cuyos pesos son menores 554.29 kg. y 3 con peso mayor, en cambio en la **muestra**$_2$ hay 6 animales con peso menor al promedio y 1 con peso mayor, lo mismo que en la **muestra**$_3$.

En la **muestra**$_2$ el promedio no es un buen representante, por lo que se buscará otro valor. El peso 200 kg. es el que más se repite (en este caso se repite 4 veces entre 7 valores que tiene la muestra), luego se puede pensar que éste es un buen representante de esa muestra.

En la **muestra**$_3$ tampoco el promedio es un buen representante, y no hay ningún valor que se repita, por lo que se determinará otro valor que resuma esa muestra. El valor de peso 230 kg., que separa a la muestra *ordenada* en dos partes iguales (es decir que hay 3 valores de la muestra menores a 230 kg. y 3 mayores) puede elegirse como tal.

La **muestra**$_1$ puede ser caracterizada por el *promedio*, en tanto que la **muestra**$_2$ y la **muestra**$_3$ pueden serlo por el *valor de variable que más se repite* y el *valor de variable que divide a la muestra en dos partes iguales*, respectivamente. Los valores descriptos son ejemplos de ***Estadísticos***.

Definición 1: Los *Estadísticos* son funciones de algunas o de todas las observaciones individuales que componen la muestra, lo que se puede expresar en forma simbólica de la siguiente manera

$$f: (X_1, X_2,...,X_n) \mapsto f(X_1, X_2,...,X_n)$$

2.2 Estadísticos de Posición

Los valores con que se han caracterizado las muestras anteriores, reciben el nombre de *estadísticos de posición* o *medidas de tendencia central*.

Definición 2: Se denominan *Estadísticos de Posición* a aquellos valores que tienden a ubicarse en el centro de la muestra ordenada.

Estos valores proporcionan una idea de los datos de la muestra y alrededor de ellos tienden a agruparse todas las observaciones de la misma.

Algunos de los estadísticos de posición son:

- **Media aritmética o promedio**: es la suma de todos los valores de la muestra dividido el tamaño de la misma. Se lo denota con $\overline{X}$ y la fórmula de cálculo es:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- **Moda**: es el valor de variable que más se repite y se denota con m.

- **Mediana**: es aquel valor que verifica que la mitad de los datos de la muestra son menores o iguales a él y la otra mitad son mayores o iguales a él. Se denota por M. Para calcularla se debe necesariamente ***ordenar*** la muestra y encontrar el valor "central" de la misma como sigue :

1. Si n es par

$$M = \frac{X_{n/2} + X_{n/2+1}}{2}$$

es decir es el promedio de los valores ubicados en el centro de la muestra ordenada.

2. Si n es impar

$$M = X_{(n+1)/2}$$

es decir es el valor ubicado en el centro de la muestra ordenada.

Si bien algunos estadísticos de posición no necesariamente asumen valores obtenidos en la muestra, todos ellos deben ser valores entre el mínimo y el máximo de la muestra estadística.

Si se dispone de los datos resumidos en una *tabla de frecuencias no agrupadas*, la media puede calcularse como

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} f_i X_i$$

Por otra parte, algunas veces no se dispone de los datos reales de la muestra y sólo se tiene acceso a la *tabla de frecuencias agrupadas* de los mismos. En esos casos se toma como representante del intervalo a la *marca de clase* del intervalo y en ese caso la fórmula para determinar la media es

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} f_i \widetilde{X}_i$$

cuyo valor no necesariamente coincide con el verdadero valor de la media.

2.2.1 Interpretación

Ejemplo 3: Usando los datos de la **muestra**$_2$ del Ejemplo 2, resultan

1) $\overline{X} = 287.14$ kg.   2) $m = 200$ kg.   3) $M = 200$ kg.

lo cual dice que

1) Los pesos de los 7 animales están alrededor de 287.14 kg.

2) El peso que más se repite es de 200 kg. (no necesariamente esto implica que la mayoría de los animales pesen 200 kg.).

3) Hay 3 animales que pesan 200 kg. o menos y 3 animales que pesan más de 200 kg.

2.2.2 Comparación entre los estadísticos de posición

Aunque desde un punto de vista puramente descriptivo las tres medidas proporcionan información complementaria, sus propiedades son muy distintas: la *media* utiliza todos los datos y es, por tanto, preferible si los datos son homogéneos; tiene el inconveniente de que es muy sensible a observaciones atípicas (un error en los datos o un valor anormal puede modificarla notablemente). Por el contrario, *la mediana* utiliza menos información que la media (sólo tiene en cuenta el orden de los datos y no su magnitud) pero, en contrapartida, no se ve alterada si una observación o una pequeña parte de las observaciones son valores atípicos. En tanto que la *moda* es el valor descriptivo más débil, ya que en algunos casos puede no existir o puede no ser única; es útil cuando la variable es de tipo cualitativo ya que es el único estadístico de posición que puede calcularse.

En general es recomendable calcular la media y la mediana ya que si hay heterogeneidad en los datos ambas medidas difieren notoriamente.

2.3 Estadísticos de Dispersión

Ejemplo 4: Suponga que los pesos de dos muestras de 7 novillos de raza Charolais tratados con la nueva dieta son:

muestra$_4$: 400 400 400 400 400 400 400
muestra$_5$: 200 300 400 400 400 500 600

Si se calculan los estadísticos de posición para cada una de ellas, se ve que son todos iguales a 400 kg., específicamente

$$\overline{X}_4 = \overline{X}_5 = m_4 = m_5 = M_4 = M_5 = 400$$

A pesar de lo anterior si se observan los pesos de los animales, se puede decir que el efecto de la nueva dieta no fue igual en las dos muestras, lo cual indica que *no son suficientes los estadísticos de posición para describir una muestra*.

Lo que se desea es medir la diferencia que se visualiza entre las dos muestras, que en realidad proviene de la variabilidad de las mismas. Para esto se calcula la diferencia entre el valor máximo de la muestra (denotado por X_{max}) y el valor mínimo (denotado por X_{min}) , es decir $X_{max}-X_{min}$, cuyo valor para cada muestra es:

400 kg. - 400 kg. = 0 kg. 600 kg. - 200 kg. = 400 kg.

Estos valores indican que en la primera muestra todos los datos son iguales, mientras que en la segunda no lo son, y ahí sí se puede observar el efecto diferente de la nueva dieta en los novillos de las dos muestras. En la primera muestra (donde todos los valores son iguales) la diferencia es cero, en tanto que, en la segunda muestra (donde no todos los valores son iguales) el valor de la diferencia es distinto de cero (positivo). Se definirán valores que tengan precisamente esas características:

* que resulten 0 cuando todos los datos son iguales;
* que sean positivos cuando hay al menos uno diferente.

Definición 3: Se denominan *Estadísticos de Dispersión* a aquellos valores que miden la variabilidad de una muestra.

Los estadísticos de dispersión más utilizados son:

- **Amplitud o Rango**: Es la diferencia entre el valor máximo y el valor mínimo observado en la muestra. Se lo denota con **w**. En fórmula se expresa

$$\mathbf{w} = X_{max} - X_{min}$$

Este estadístico no utiliza toda la información de la muestra (sólo sus extremos), por ello se proponen otras medidas de dispersión que sí la tienen en cuenta. Una de ellas surge en considerando la diferencia de las observaciones con respecto a la media muestral: $\left(X_i - \overline{X}\right)$. Sin embargo, como ésta es una medida para cada observación y se quiere estudiar la variabilidad de todos los valores de la muestra, se deberían sumar todas estas diferencias (en símbolos, $\sum_{i=1}^{n}\left(X_i - \overline{X}\right)$). Se puede comprobar que esta suma es *siempre* cero cualesquiera

sean los datos de la muestra, motivo por el cual no puede ser una medida de variabilidad. Entonces basados en esta idea se define el siguiente estadístico de dispersión:

- **Varianza:** Es un valor que mide cuanto se desvían en promedio los datos de la media muestral. Se lo denota con S^2 y en notación matemática se lo expresa así

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$$

Observe que S^2 vale cero cuando los datos son todos iguales y es mayor que cero cuando al menos uno es diferente, con lo que se logra lo que se espera de una medida de variabilidad, utilizando además todos los valores de la muestra.

Dado que este estadístico tiene las unidades de los datos elevadas al cuadrado, se define otro estadístico que tiene la misma magnitud que los datos.

Si se dispone de los datos resumidos en una *tabla de frecuencias no agrupadas*, la varianza puede calcularse como

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i\left(X_i - \overline{X}\right)^2$$

Por otra parte, si sólo se tiene acceso a la *tabla de frecuencias agrupadas* de los datos la varianza (usando la marca de clase) se calcula como

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i\left(\widetilde{X}_i - \overline{X}\right)^2$$

- **Desviación Estándar:** Este estadístico mide, al igual que la varianza, cuanto se desvían en promedio los datos de la media muestral, pero con la misma magnitud que los datos. Se lo expresa de la siguiente manera:

$$S = +\sqrt{S^2}$$

- **Coeficiente de Variación:** Es el estadístico de dispersión que mide la variabilidad de la muestra independientemente de la magnitud de la media, o sea determina la variabilidad en unidades de la media. Se lo denota con **CV** y se lo calcula como

$$CV = \frac{S}{\overline{X}} \bullet 100$$

El coeficiente de variación es un número sin unidades de medida, generalmente expresado en porcentaje. Esta ausencia de unidades es de gran utilidad cuando se desea comparar la variabilidad de dos o más muestras.

2.3.1 Interpretación

Retomando el Ejemplo 4 se puede observar, a simple vista que, para la primera muestra tanto S_4^2 como S_4 son cero, en tanto que para la segunda muestra ambos valores resultan mayores a cero:

$$S_5^2 = \frac{(200-400)^2 + (300-400)^2 + 3.(400-400)^2 + (500-400)^2 + (600-400)^2}{6} = 16666.67\text{kg}^2$$

y $S_5 = 129.10$ kg.

La interpretación de los valores numéricos de los estadísticos obtenidos en la **muestra**$_5$ de novillos de la raza Charolais a los que se les suministró la nueva dieta es:

- $w = 400$ kg. significa que *la diferencia entre los pesos del animal más pesado y más liviano es 400 kg.*
- $S_5 = 129.10$ kg. *significa que los pesos de los 7 animales se desvían aproximadamente 129.10 kg. del peso promedio*.
- $CV_5 = 32.275$ % significa que *la variabilidad de los pesos relativa al peso promedio es aproximadamente del 32%*.

Para comprender cuan útil es el coeficiente de variación se da el siguiente

Ejemplo 5: Se sospecha que el peso de las ratas es más variable que el peso de los elefantes, para lo cual se determinó el peso medio y la desviación estándar de los pesos de los animales en estudio, los cuales resultaron:

$\overline{X}_E = 10406.4$ kg. $S_E = 557.68$ kg.

$\overline{X}_R = 0.46$ kg. $S_R = 0.07$ kg.

Si se tiene en cuenta el valor de las desviaciones estándar puede parecer que los pesos de los elefantes varían mucho más que los pesos de las ratas. Sin embargo al calcular el coeficiente de variación para cada muestra se observa que

$CV_E = 5.36$ % y $CV_R = 15.22$ %

lo cual indica que en realidad varía más el peso de las ratas que el peso de los elefantes, es decir que la muestra con los datos de los pesos de los elefantes es más uniforme que la de los pesos de las ratas.

2.3.2 Comparación entre los estadísticos de dispersión

El *rango* brinda una rápida visualización de la variabilidad de la muestra, aunque sólo utiliza los valores extremos de la misma. Por su parte la *varianza* utiliza toda la información de la muestra pero no posee las unidades de medida de los datos, condición que sí es verificada por la *desviación estándar*. Sin embargo, si el interés es comparar la variabilidad de dos o más muestras, la desviación estándar no es aconsejable pues su valor está en la magnitud de los datos; en ese caso el *coeficiente de variación* es el apropiado.

2.4 Diagrama de Caja

Si bien la media y la desviación estándar son las medidas descriptivas más comunes, existen otras que proporcionan información adicional acerca de las características de un conjunto de datos. Por ejemplo:

- **Primer Cuartil:** es la mediana de los valores inferiores o iguales a la mediana de la muestra original. Este valor verifica que la cuarta parte de los datos de la muestra son menores o iguales a él y las tres cuartas partes de los datos son mayores o iguales a él. Para calcularlo se debe necesariamente ***ordenar*** la muestra. Se denota por Q_1.
- **Segundo Cuartil:** es la Mediana y se denota por Q_2.
- **Tercer Cuartil:** es la mediana de los valores superiores o iguales a la mediana de la muestra original. En este caso las tres cuartas partes de los datos de la muestra son menores o iguales a él y la cuarta parte de los datos son mayores o iguales a él. También para calcularlo se debe necesariamente ***ordenar*** la muestra . Se denota por Q_3.

A partir de estas medidas es posible construir el gráfico de caja, el cual proporciona información útil para un análisis descriptivo integral del conjunto de datos (Freund, J. y Manning Smith, R. 1989). La forma más simple de construir éste es a través de los siguientes 5 valores

1. El mínimo
2. El primer cuartil
3. El segundo cuartil
4. El tercer cuartil
5. El máximo

Para los datos del Ejemplo 1, el diagrama de caja es

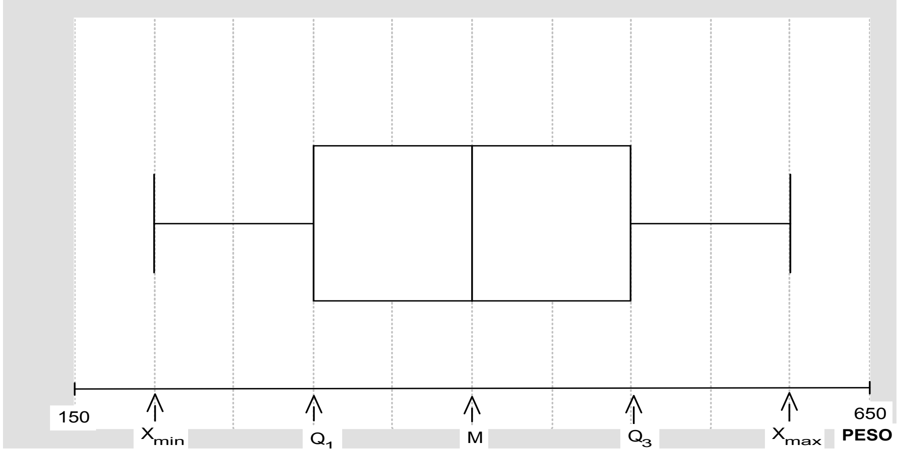


Gráfico 1: Distribución de los novillos según el peso

2.5 Descripción conjunta de dos variables

La medida de asociación lineal más simple entre n pares de observaciones aleatorias $(X_1,Y_1),\ldots,(X_n,Y_n)$ es la covarianza definida por:

$$\mathrm{Cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

El inconveniente de la covarianza como medida de asociación es su dependencia de las unidades de medida de las variables. Por ejemplo, la covarianza entre la altura y el peso es 200 (cuando las variables se miden en centímetros y gramos respectivamente) mientras que la covarianza resulta 0.002 (cuando se expresa la altura en metros y el peso en kilogramos). Éstos valores tan diferentes llevarían a pensar que las observaciones medidas en ciertas magnitudes tienen una asociación lineal diferente que cuando se expresan en otras magnitudes, lo cual es incorrecto. Para resolver esta dificultad se construye una medida adimensional, dividiendo la covarianza por un término con sus mismas dimensiones, dando lugar al siguiente estadístico:

- **Coeficiente de Correlación:** mide el grado de asociación lineal entre n valores de las variables X e Y y se define por:

$$r = \frac{\mathrm{Cov}(X,Y)}{S_X S_Y}$$

donde S_X y S_Y son las desviaciones estándares de las variables X e Y respectivamente. Se puede ver que:

1. El coeficiente de correlación es adimensional: su valor no varía si se multiplica cualquiera de las variables por una constante positiva.
2. Si existe relación lineal exacta entre las variables (todos los puntos están en la línea), el coeficiente de correlación es igual a 1 o -1.
3. Si no existe relación lineal exacta se cumple que $-1 < r < 1$.

Ejercicios de Aplicación

1.

Los datos de la tabla corresponden a dos muestras de salarios (en $) de los empleados en dos establecimientos ganaderos:

a) Calcular para cada muestra los estadísticos de posición y compararlos.

b) Según lo obtenido en a), ¿reflejan los estadísticos de posición la situación salarial de ambos establecimientos?. Justificar.

c) Calcular los estadísticos que crea conveniente para reflejar lo observado en b).

d) Representar gráficamente las muestras estadísticas y discutir la relación entre los valores de los estadísticos y la forma de los gráficos.

Establecimiento A	Establecimiento B
200	300
200	300
200	300
300	300
300	300
300	300
300	300
900	400

2.

a) Dar un ejemplo de:
i) una muestra de tamaño 5, con media 5 y dispersión nula;
ii) dos muestras con igual media y distinta varianza;
iii) una muestra de 7 datos de manera que el estadístico más representativo sea la moda.

b) Si en una granja hay pollos cuyos pesos varían entre 2.8 kg. y 3.8 kg.:
i) ¿Puede el peso medio tomar los valores 2.1 kg. ó 4.2 kg. ?. ¿Por qué?.
ii) ¿Es posible que la varianza sea nula?.

3.

En cada uno de los siguientes ejemplos decidir si es posible que los estadísticos tomen los valores que se informan:

a) En 5 vacunos Ayrshire se encontró que el mínimo porcentaje de grasa en la leche fue del 4%, el promedio fue de 14% y el rango de 14%.

b) Para otro grupo de 5 vacunos de la misma raza mencionada en el punto anterior, se encontró el mismo valor mínimo de porcentaje de grasa, pero en cambio el rango fue de 10% y el valor más frecuente fue el 14%.

c) En una quinta se seleccionaron 8 repollos y se les contó el número de larvas de cierto insecto y además se determinó su peso. Para la primera variable el rango fue de 6 larvas, mientras que en la segunda variable el rango fue de 0.5 kg.. Ambas muestras estadísticas carecieron de moda.

d) En una muestra de huevos de ponedoras Leghorn blanca el rango de pesos fue de 9 gr., con un valor más frecuente de peso de 45 gr. . El huevo más pesado presentó un peso de 45 gr..

e) Siete ponedoras de la raza mencionada en el punto anterior fueron seleccionadas para ser pesadas. Los valores de peso encontrados oscilaron entre 1.4 kg. y 2.4 kg.. El peso más frecuente es 1.4 kg..

4.

En el Ejercicio 6 del Capítulo 1 se presentaron datos sobre la distribución del área sembrada entre los años 83 y 87. ¿Para que cereal hubo mayor variabilidad de la distribución del área sembrada ?.

5.

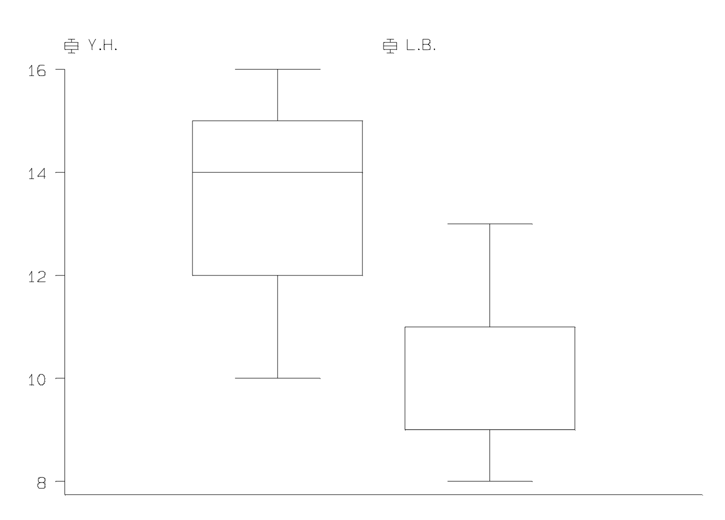
En el Ejercicio 4 del Capítulo 1 se presentaron los números de lechones por año para cerdas de la raza Yorkshire holandés. Repitiendo la misma experiencia pero para 30 cerdas de la raza Landrase belga, se obtuvo el siguiente conjunto de datos:

Número de lechones por año	Número de cerdas
8	7
9	9
10	4
11	6
12	2
13	2

a) Calcular los estadísticos que crea conveniente para mostrar las diferencias en el comportamiento de la variable “número de lechones por año” para ambos grupos de cerdas.

b) Realizar diagramas de barra para comparar el comportamiento de ambas muestras estadísticas.

c) En base a los diagramas de cajas para los dos conjuntos de datos presentados, discutir lo observado en los gráficos y relacionar con los estadísticos obtenidos en el inciso a).



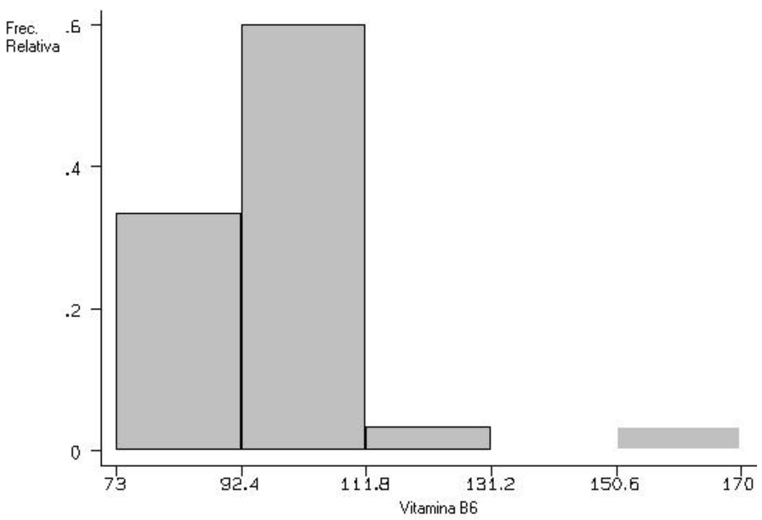
Diagramas de Caja para el "número de lechones" en dos razas de cerdas: Yorkshire Holandés y Landrase Belga

6.

Se seleccionaron un total de 30 muestras de 100 gramos de quesos Sbrinz determinándose el contenido de vitamina B6 (en mcg.), obteniéndose la siguiente muestra estadística:

93	170	92	96	84	73
96	100	97	95	80	95
95	90	120	94	97	96
98	86	78	94	96	95
97	87	76	92	98	96

Además se construyó el histograma para las frecuencias relativas de los contenidos de vitamina B6 que se presenta a continuación:



Histograma para las frecuencias relativas de los contenidos de vitamina B6 en muestras de queso Sbrinz

a) Calcular los estadísticos que crea conveniente para reflejar el comportamiento de los datos mostrado en el histograma.

b) A continuación se muestra un histograma donde se grafican los contenidos de vitamina B6 (en mcg.) para 25 muestras de Queso Fundido. ¿Qué diferencias

substanciales encuentra entre este nuevo gráfico y el histograma para las muestras de queso Sbrinz?. Intuitivamente ¿cómo se reflejaría este cambio del comportamiento de los contenidos de vitamina B6 en los estadísticos de posición y dispersión?.



Histograma para las frecuencias relativas de los contenidos de vitamina B6 en muestras de queso Fundido.

7.

A fin de comparar los pesos de los huevos de tres líneas de ponedoras Leghorn blanca, A, B y C, se tomaron 5 huevos de cada una de tales líneas, observándose los siguientes valores (en gr.):

Línea A	46.6	47.1	48	54.3	45.7
Línea B	45.3	45.2	46.1	44.2	43.2
Línea C	54	52.1	53.6	52.6	56

a) ¿ Cuál es la variable en estudio y a qué tipo corresponde?.

b) Realizar un estudio comparativo entre las tres muestras estadísticas.

8.

Un equipo de investigadores intenta establecer alguna relación entre el consumo de agua y el alimento de pollas Leghorn estándar colocadas en jaulas. Para ello se midió durante varios días el consumo diario de alimento cada 100 pollas (en kilogramos) y el consumo diario de agua cada 100 pollas (en libras). Los datos siguientes corresponden a la muestra estadística:

Consumo de Alimento	11.8	11.6	11	10	8.7	7	4.8
Consumo de Agua	33.9	35.7	38.6	44	55.4	73.5	89.2

A partir del siguiente resumen de los principales estadísticos para ambas muestras realizar un informe, que usted crea conveniente, del comportamiento de los datos:

Estadísticos	**Consumo de Alimento**	**Consumo de Agua**
Media	9.27	52.9
Mediana	10	44
Varianza	6.8	448.14
Mínimo	4.8	33.9
Máximo	11.8	89.2

3 Probabilidades

Objetivos:

- Reconocer los ámbitos de aplicación de las distintas definiciones de la probabilidad en distintos contextos.
- Distinguir entre sucesos excluyentes y sucesos independientes.

3.1 Introducción

Hasta ahora, sólo se han planteado problemas tales como: un investigador intenta imponer
* un nuevo fármaco;
* una nueva dieta;
* una nueva variedad.

Para resolver estos problemas o para tomar una decisión, por ejemplo sobre si la nueva dieta es mejor que la utilizada, se ha tomado una muestra y se ha realizado un análisis descriptivo de la misma mediante gráficos, tablas y/o estadísticos. Pero también, se debe tener en claro que el problema está planteado con fines de conocer características de la población, las cuales pueden ser cuantificadas con valores típicos que reciben el nombre de *características numéricas de la variable o de la población estadística*. Sin embargo, se debe observar también que es imposible o poco práctico (por razones de tiempo o de costo) a menos que la población en estudio sea pequeña, inspeccionar todas las unidades experimentales de la misma por lo que se recurre al estudio de una muestra extraída de esa población.

Definición 1: Una característica descriptiva global de una población estadística se llama *característica numérica* de la variable.

Los valores típicos de la población estadística pueden ser también denominados parámetros, pero aquí se reservará este término para las constantes que identifican un modelo probabilístico, como se mostrará en los Capítulos 4 y 5.

Definición 2: Una característica descriptiva global de una muestra estadística se llama *estadístico*.

A continuación se muestran los estadísticos y las características numéricas más utilizados para describir una muestra y una población respectivamente.

MUESTRA (Estadísticos)	**POBLACIÓN (Características Numéricas de la Variable)**
$\overline{X}$	$E(X)$
S^2	$Var(X)$
S	$Des(X)$

Raramente se utilizan los estadísticos sólo como resumen descriptivo, se usan más bien como indicadores de alguna característica de la población de la cual fue extraída la muestra. Así los valores muestrales $\overline{X}$ y S^2 son de escaso interés a menos que proporcionen alguna información acerca del *Promedio poblacional* y de la *Varianza poblacional.*

Las características numéricas de la variable en estudio mencionadas son generalmente desconocidas, pues su valor sólo puede obtenerse examinando a toda la población, lo cual es prácticamente imposible. Por ello es que se trata de *estimarlos* con los valores muestrales.

Ahora bien, aunque se trate de tomar la "mejor muestra", es decir, la más representativa, nunca se podrán obtener las características poblacionales con un 100% de seguridad, por no conocer la información de toda la población.

Estimar características de una población a través de características de una muestra, recibe el nombre de **Inferencia Estadística.** Ese hecho tiene un costo, que es el de tener conocimiento sobre la población con un cierto margen de error, el cual es medido a través de una *probabilidad*. Luego se puede dividir a la estadística en dos grandes ramas vinculadas por medio de la probabilidad y modelos probabilísticos.

- Estadística Descriptiva **- Estadística Inferencial**

No siempre la finalidad al obtener una muestra es tratar de hacer inferencia. Hay situaciones en donde sólo importa describir lo que está ocurriendo en la muestra para lo que se utilizan las herramientas de la Estadística Descriptiva, tales como gráficos, tablas y estadísticos.

3.2 Algunas definiciones básicas

Se verán a continuación ciertas nociones necesarias para estudiar el concepto de probabilidad.

- Los fenómenos *determinísticos* son aquellos cuyos resultados se pueden predecir. Algunos fenómenos físicos son de este tipo, por ejemplo si se arroja una roca hacia arriba por la ley de la gravedad va a caer y por tanto no hay nada de incierto en el resultado de este experimento.

- Los fenómenos *aleatorios* son aquellos cuyos resultados no se pueden predecir. La estadística está basada justamente en experimentos de este tipo.

Definición 3: Un *experimento aleatorio*, es un proceso cuyo resultado no se puede predecir. Se lo denota con ***E***.

También se puede decir que los experimentos aleatorios son fenómenos empíricos que se caracterizan por una propiedad fundamental y propia: su observación repetida en condiciones constantes, no produce el mismo resultado porque no existe regularidad determinística sino regularidad estadística o aleatoria.

Ejemplo 1:

1. Arrojar una moneda. 2. Arrojar un dado
3. Suministrar una dieta. 4. Sembrar una variedad.

Definición 4: El *espacio muestral* de un experimento aleatorio es el conjunto de todos los resultados posibles del mismo. Se lo denota con ***S***.

Ejemplo 2: El espacio muestral correspondiente a cada uno de los experimentos aleatorios indicados en los puntos 1 y 2 del Ejemplo 1 son

1.***S*** = {C, X} 2. ***S*** = {1,2,3,4,5,6}

Sobre un experimento aleatorio se pueden definir diferentes sucesos aleatorios:

Definición 5: Un *suceso aleatorio* es un conjunto formado por algunos o todos los resultados posibles de un experimento aleatorio. En términos de conjunto se puede decir que un suceso es un subconjunto del espacio muestral. En general se denotan con letras mayúsculas ***A***, ***B***, ***C***.

Definición 6: Un *suceso elemental* es aquel suceso que tiene un solo resultado posible (conjunto con un solo elemento).

Definición 7: El *suceso imposible* es el que no ocurre cuando se efectúa el experimento. Se lo denota con $\varnothing$.

Definición 8: El *suceso seguro* es el que ocurre siempre. Coincide con el espacio muestral.

Ejemplo 3: Sea el experimento ***E***: “arrojar una moneda dos veces ”

- El espacio muestral asociado a él es: ***S*** = {(C,C); (C,X); (X,C); (X,X)}

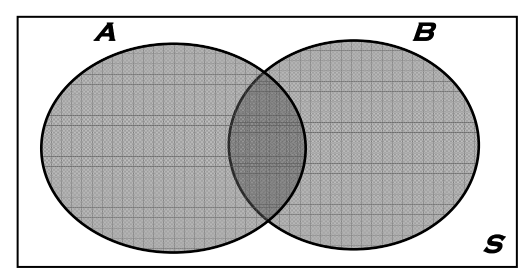
Sobre este experimento se definen los sucesos aleatorios

- ***A***: “se obtiene al menos una cara”, luego por extensión ***A*** = {(C,C); (X,C); (C,X)}
- ***B***: “se obtienen dos caras”, luego ***B*** ={(C,C)}
- ***C***: "se obtienen tres caras”, luego ***C*** = {(C,C,C)} y como éste no es un subconjunto de ***S*** resulta ser un suceso imposible.

3.3 Relaciones entre sucesos

Sean ***A*** y ***B*** dos sucesos

a) El *suceso suma* denotado por ***A***+***B***, es el suceso que ocurre si ***A*** o ***B*** o ambos ocurren. Este suceso puede ser visualizado utilizando los diagramas de Venn, como se muestra a continuación, para las dos situaciones que se pueden presentar.



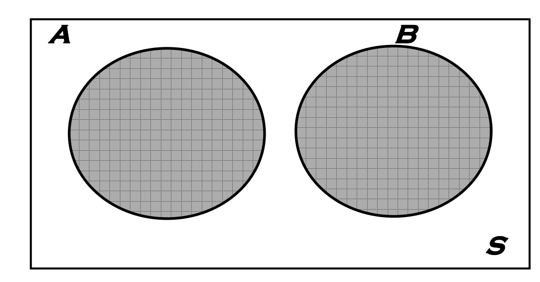


Gráfico 1: Suceso suma

b) El *suceso producto*, denotado por ***A.B***, es el suceso que ocurre si ***A*** y ***B*** ocurren simultáneamente, el que puede ser visualizado utilizando los diagramas de Venn, como sigue

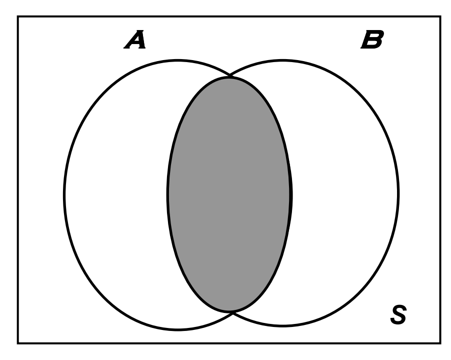


Gráfico 2: Suceso producto

c) *El suceso complemento* de ***A***, denotado por A^C, es el suceso que ocurre cuando el suceso ***A*** no ocurre.

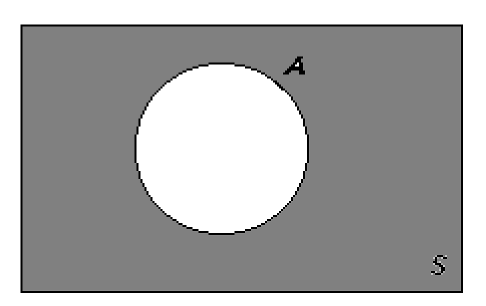


Gráfico 3: Suceso complemento

Ejemplo 4: Sea el experimento ***E***: "arrojar un dado" y ***S*** = { 1,2,3,4,5,6 }

Se definen sobre el experimento los siguiente sucesos

A: "sale un número par", luego por extensión ***A*** = { 2, 4, 6 }
B: "sale un número impar", luego ***B*** = { 1, 3, 5 }
C: "sale un número menor que 4", luego ***C*** = { 1, 2, 3 }

Así los sucesos ***A***+***C***, ***A***.***B***, ***B***.***C*** y C^C resultan:

1. ***A***+***C*** = {1, 2, 3, 4, 6}   2. ***A***.***B*** = ∅   3. ***B***.***C*** = {1, 3}   4. C^C = {4 ,5, 6}

Definición 9: Dos sucesos son *excluyentes* si no pueden ocurrir simultáneamente. En símbolos ***A*** y ***B*** son excluyentes si y sólo si ***A***.***B*** = ∅.

En el Ejemplo 4, al arrojar el dado no puede salir simultáneamente un número par y un número impar, o sea, se verifica que ***A***.***B*** = ∅, es decir los sucesos ***A*** y ***B*** son excluyentes.

3.4 Definición clásica de probabilidad

Para introducir la noción de probabilidad se utiliza el siguiente

Ejemplo 5: En una urna hay 6 fichas idénticas y bien mezcladas 5 de color y 1 blanca. Si se realiza el siguiente experimento aleatorio

E: "extraer una ficha de la urna”

el espacio muestral asociado es

$$S = \{b, c_1, c_2, c_3, c_4, c_5\} .$$

Teniendo en cuenta el espacio muestral, ¿es de esperar que la ficha blanca tenga la misma posibilidad de ser extraída que una ficha de color?. La respuesta es no, ¿cómo cuantificar esa posibilidad?. Para ello se definen los siguientes sucesos aleatorios

A: “salga una ficha de color” y ***B***: “salga la ficha blanca”.

Las posibilidades de obtener una ficha de color son mayores que las de obtener una ficha blanca. Exactamente son 5/6 y 1/6 respectivamente, donde el número 6 indica la cantidad total de fichas de la urna, es decir el número total de resultados posibles del experimento (cantidad de elementos del espacio muestral) y el número 5 indica la cantidad de fichas de color, es decir la cantidad de resultados favorables al suceso ***A*** y el número 1 la cantidad de fichas blancas, es decir la cantidad de resultados favorables al suceso ***B***. A los valores 5/6 y 1/6 se los llama *probabilidad*.

Definición 10: La *probabilidad* de un suceso ***A*** es el cociente entre el número de resultados favorables al suceso ***A*** y el número de resultados posibles del experimento (**m** y **n** respectivamente). En símbolos

$$P(A) = \frac{m}{n}$$

Es importante notar que para aplicar esta definición deben verificarse dos supuestos muy importantes

1. El número de elementos del espacio muestral debe ser *finito*.
2. Todos los resultados del experimento aleatorio deben ser *igualmente posibles* (equiprobables).

Por ello, en el Ejemplo, se resalta el hecho que las fichas "son idénticas" y "están bien mezcladas".

Ejemplo 6: Suponga que de un mazo de 50 cartas un jugador recibe una carta. ¿Cuál es la probabilidad de que el jugador

a) reciba un rey? b) reciba una copa?
c) reciba una carta? d) no reciba ninguna carta?

Solución:

- Para resolver este problema se debe observar primero si se verifican los dos supuestos indicados anteriormente. En este caso, el número de elementos del espacio muestral es finito $\mathbf{n}=50$ y, bajo el supuesto de que el mazo esté bien barajado, cada carta tiene la misma posibilidad de ser extraída, esto es $P(\text{obtener una carta cualquiera})=1/50$.

- Si se denotan ***A***, ***B***, ***C*** y ***D*** los sucesos definidos en cada uno de los incisos anteriores, se tiene:

$$P(\boldsymbol{A}) = 4/50,\ P(\boldsymbol{B}) = 12/50\ ,\ P(\boldsymbol{C}) = P(\boldsymbol{S}) = 50/50=1\ \text{ y } P(\boldsymbol{D}) = P(\emptyset) = 0/50=0$$

En general se verifica:

1. $0 \leq P(\boldsymbol{A}) \leq 1$, para cualquier suceso ***A***.

2. $P(\boldsymbol{S}) = 1$.

3. $P(\emptyset) = 0$.

En base a las Definiciones 9 y 10 se deducen las siguientes consecuencias:

- ***A.B*** = $\emptyset$ sí y sólo sí $P(\boldsymbol{A.B}) = 0$.
- Si $P(\boldsymbol{A.B}) \neq 0$ entonces ***A*** y ***B*** son *no excluyentes*.

3.5 Definición estadística de probabilidad

A continuación se presenta la idea de *regularidad estadística*, nombre que se admite como adecuado para indicar el siguiente hecho: si un fenómeno se repite en las mismas condiciones un número considerable de veces $n_1, n_2, n_3,...$ (con $n_1 > n_2 > n_3 >...$), se determina para el suceso de interés ***A*** la frecuencia relativa (f_{rA}) y si se observa que dichas frecuencias relativas tienden a estabilizarse alrededor de un número, se dice que el fenómeno se comporta con regularidad estadística. La definición estadística de probabilidad de un suceso está basada en esto.

Para fijar esta idea se presentan los siguientes ejemplos, a partir de los cuales se deducirá la definición estadística de probabilidad.

Ejemplo 7: Sea el experimento aleatorio ***E***: "arrojar una moneda balanceada" y sean $n_1=1$, $n_2=10$, $n_3=100,...$ las veces que se repite el experimento. El suceso de interés en este caso es ***A***: "obtener una cara". Las frecuencias relativas de cara obtenidas se representan en el Gráfico 1. Si se observa este gráfico se ve que las frecuencias relativas correspondientes al suceso ***A*** tienden a estabilizarse alrededor de un número fijo 1/2 .



Gráfico 4: Proporción de caras según el Número de tiradas.

Ejemplo 8: Los siguientes datos son obtenidos de una publicación de cifras demográficas relativas a nacimientos de bebés clasificados por sexo, se tomaron muestras de tamaño 10, 100, 1000, 10000, 100000 nacimientos. La frecuencia para cada sexo en cada muestra se indican en la Tabla 1. Observando la columna de f_{rV} se nota fácilmente que a medida que aumenta el número de nacimientos, la proporción de varones nacidos tiende a estabilizarse alrededor de 0.51. Para simplificar este ejemplo no se consideraron los nacimientos múltiples y además sólo se tomaron los nacidos vivos.

Tabla 1: Frecuencias por sexo

Tamaño de muestra	f_V	f_M	f_{rV}	F_{rM}
10	6	4	0.60000	0.40000
100	55	45	0.55000	0.45000
1000	526	474	0.52600	0.47400
10000	5187	4813	0.51870	0.48130
100000	51005	48995	0.51005	0.48995

Todo lo anterior conduce a la siguiente definición de probabilidad en términos de la frecuencia relativa, la cual puede considerarse la *Definición Estadística o Frecuencial de Probabilidad.*

Definición 11: A medida que aumenta el número n de repeticiones de un experimento ***E***, ejecutado en las mismas condiciones, la frecuencia relativa correspondiente a un suceso cualquiera ***A*** tiende a estabilizarse en un número. Ese número es el que se llama *Probabilidad* y puede ser expresado como

$$\lim_{n \to \infty} f_{rA} = P(A)$$

o sea cuando el tamaño de la muestra es grande se tiene que

$$f_{rA} \cong P(\boldsymbol{A})$$

Si se comparan las Definiciones 10 y 11 se puede deducir que, para obtener la *probabilidad de un suceso en base a la definición clásica no es necesario* realizar el experimento, en tanto que si se la desea obtener en base a la *definición estadística se lo debe* realizar pues la misma está en función de la frecuencia relativa.

Otra manera de presentar la noción de probabilidad está basada en el método axiomático (Meyer, P. 1992). Este abordaje requiere la realización del experimento y puede aplicarse sobre espacios muestrales infinitos Esta forma de definir la probabilidad no será presentada en este texto.

3.6 Probabilidad de algunos sucesos importantes

En la Sección 3.3 se definieron algunas relaciones entre sucesos. Se muestra a continuación la forma de calcular la probabilidad de esos sucesos.

3.6.1 Probabilidad del Suceso Suma

La probabilidad de la suma de dos sucesos ***A*** y ***B*** es

$$P(\boldsymbol{A}+\boldsymbol{B}) = P(\boldsymbol{A}) + P(\boldsymbol{B}) - P(\boldsymbol{A}.\ \boldsymbol{B})$$

Cuando ***A*** y ***B*** son *excluyentes*, $P(\boldsymbol{A}.\ \boldsymbol{B}) = 0$ y entonces la probabilidad del suceso suma es

$$P(\boldsymbol{A}+\boldsymbol{B}) = P(\boldsymbol{A}) + P(\boldsymbol{B})$$

Este resultado se puede demostrar fácilmente utilizando algunos conceptos de la teoría de conjunto.

Ejemplo 9: Hallar la probabilidad de sacar un rey o un corazón de un mazo de 52 cartas bien mezcladas.

Solución:

E: "extraer una carta al azar" ***S*** = {x / x es una carta del mazo}

A = {sacar un rey} ***B*** = {sacar un corazón} y ***A***. ***B*** = {sacar un rey de corazón}

A + ***B*** = {sacar un rey o un corazón}

Ahora se dispone de todos los datos necesarios para dar la solución del problema planteado, que en términos de probabilidad es

$$P(\boldsymbol{A}+\boldsymbol{B}) = P(\boldsymbol{A}) + P(\boldsymbol{B}) - P(\boldsymbol{A}.\ \boldsymbol{B}) = 4/52 + 13/52 - 1/52 = 16/52 \cong 0.31$$

Esto significa que *si se repite muchas veces el experimento **E** en aproximadamente el 31% de los casos sale un rey o un corazón*.

Ejemplo 10: Un tirador dispara a un blanco dividido en tres zonas. La probabilidad de impactar en la primera zona es 0.45 y en la segunda es 0.35. Hallar la probabilidad de que pegue en la primera o en la segunda zona.

Solución:

E: " tirar al blanco" ***S*** = $\{z_1, z_2, z_3\}$

Es este caso se definen los sucesos

A = {impactar en la zona 1} ***B*** = {impactar en la zona 2}

Es importante notar que los sucesos ***A*** y ***B*** son excluyentes, entonces

$$P(\boldsymbol{A}+\boldsymbol{B}) = P(\boldsymbol{A}) + P(\boldsymbol{B}) = 0.35 + 0.45 = 0.80$$

3.6.2 Probabilidad del suceso complemento

La probabilidad del suceso $\boldsymbol{A}^C$ es igual a 1 menos la probabilidad del suceso ***A***. En símbolos

$$P(\boldsymbol{A}^C) = 1 - P(\boldsymbol{A})$$

Para verificar esta igualdad se debe tener en cuenta que el espacio muestral se puede pensar como la unión de dos subconjuntos excluyentes, es decir: $\boldsymbol{S} = \boldsymbol{A} \cup \boldsymbol{A}^C$ (o equivalentemente

$\boldsymbol{S} = \boldsymbol{A} + \boldsymbol{A}^C$). Aplicando el operador probabilidad a ambos miembros y recordando que los sucesos $\boldsymbol{A}$ y $\boldsymbol{A}^C$ son excluyentes se tiene que

$$P(\boldsymbol{S}) = P(\boldsymbol{A}) + P(\boldsymbol{A}^C)$$

y dado que $P(\boldsymbol{S})=1$ se puede escribir

$$1 = P(\boldsymbol{A}) + P(\boldsymbol{A}^C)$$

y despejando se obtiene la expresión inicial.

Ejemplo 11: La probabilidad de que un equino tenga tumores cutáneos es 0.70. ¿Cuál es la probabilidad de que un equino elegido al azar no tenga tumores cutáneos?.

Para determinar esta probabilidad se aplica el resultado anterior. Luego se tiene

$$P(\boldsymbol{A}^C) = 1 - P(\boldsymbol{A}) = 1 - 0.70 = 0.30$$

3.6.3 Probabilidad del suceso producto

Antes de determinar la probabilidad del suceso producto, se presentan los conceptos de sucesos independientes y sucesos dependientes.

Definición 12: Dos sucesos ***A*** y ***B*** se dicen *estadísticamente independientes* cuando la probabilidad de que ocurra uno de ellos (***A***) no depende de la ocurrencia o no del otro (***B***).

En términos probabilísticos esta definición se puede escribir de la siguiente manera

$$P(\boldsymbol{A} / \boldsymbol{B}) = P(\boldsymbol{A})$$

o análogamente

$$P(\boldsymbol{B} / \boldsymbol{A}) = P(\boldsymbol{B})$$

Definición 13: Dos sucesos se dicen *estadísticamente dependientes* cuando la probabilidad de que ocurra uno de ellos (***A***) se ve afectada por la ocurrencia de otro (***B***). En símbolos

$$P(\boldsymbol{A} / \boldsymbol{B}) \neq P(\boldsymbol{A}) \quad \text{o} \quad P(\boldsymbol{B} / \boldsymbol{A}) \neq P(\boldsymbol{B})$$

La expresión $P(\boldsymbol{A} / \boldsymbol{B})$ indica la probabilidad de un suceso ***A*** condicionado a la ocurrencia de otro suceso ***B***.

Definición 14: La *Probabilidad Condicional* de un suceso ***A*** dado otro suceso ***B*** es igual al cociente entre la probabilidad del suceso producto ***A.B*** y la probabilidad del suceso ***B***. En símbolos

$$P(\boldsymbol{A} / \boldsymbol{B}) = P(\boldsymbol{A}.\boldsymbol{B}) / P(\boldsymbol{B}) \qquad \text{si } P(\boldsymbol{B}) \neq 0$$

Análogamente

$$P(\boldsymbol{B} / \boldsymbol{A}) = P(\boldsymbol{A}.\boldsymbol{B}) / P(\boldsymbol{A}) \qquad \text{si } P(\boldsymbol{A}) \neq 0$$

Para comprender esta definición se presenta el siguiente ejemplo

Ejemplo 12: La Tabla 2 reproduce los datos del Problema 1.6 que trata sobre tumores cutáneos en equinos según el color del pelaje.

Tabla 5[1]: Distribución de frecuencias según el color y presencia del tumor.

PRESENCIA DE TUMOR COLOR	Si	No	TOTAL
Alazán	220	80	300
Zaino	135	115	250
Tordillo	415	35	450
TOTAL	770	230	1000

Sea entonces el experimento: ***E***: “Extraer un animal al azar" .

a) ¿Cuál es la probabilidad de que el animal elegido sea de color zaino?.

b) ¿Cuál es la probabilidad de que el animal sea de color zaino dado que tiene tumor cutáneo?.

c) ¿ Los sucesos “color zaino” y “presencia de tumor cutáneo” son sucesos independientes ?.

Solución:

a) En primer lugar se debe definir el suceso ***A*** = {animal de color zaino} entonces

$$P(\boldsymbol{A}) = n_A / n = 250/1000$$

Lo que indica que aproximadamente el 25% de los animales son de color zaino.

b) Interesa calcular la probabilidad condicional $P(\boldsymbol{A} / \boldsymbol{B})$ donde ***B*** = {animal que presenta tumor}. Para calcular esta probabilidad se puede utilizar la definición clásica. La información adicional “presenta tumor cutáneo" reduce el espacio muestral ***S*** al subconjunto de animales con tumor cutáneo, de allí que los casos posibles corresponden a la cantidad de animales con tumores cutáneos (en este caso 770), mientras que los casos favorables corresponden a los animales que verificaron las dos condiciones "color zaino y presenta tumor ". Entonces

$$P(\boldsymbol{A / B}) = \frac{n_{A.B}}{n_B} = \frac{135}{770} \cong 0.18$$

El mismo resultado se obtiene si se aplica la Definición 14

$$P(\boldsymbol{A / B}) = \frac{P(\boldsymbol{A.B})}{P(\boldsymbol{B})} = \frac{135/1000}{770/1000} = \frac{135}{770} \cong 0.18$$

Dada la equivalencia entre ambos modos de calcular la probabilidad condicional, cuando se cuenta con tablas de contingencia es recomendable usar la definición clásica (reduciendo el espacio muestral).

c) La independencia se verifica si $P(\boldsymbol{A} / \boldsymbol{B}) = P(\boldsymbol{A})$. En este caso

[1] *Coincide con la Tabla 5 del Capítulo 1.*

$$P(\boldsymbol{A} / \boldsymbol{B}) = 135/770 \cong 0.18 \qquad \text{y} \qquad P(\boldsymbol{A}) = 250/1000 = 0.25$$

como $P(\boldsymbol{A} / \boldsymbol{B}) \neq P(\boldsymbol{A})$ se concluye que los sucesos "color zaino" y "presencia de tumor cutáneo" no son sucesos independientes.

Se supone que $n=1000$ es un tamaño suficientemente grande como para verificarse la aproximación adecuada entre las frecuencias relativas y las probabilidades respectivas.

Una vez resuelto el problema de cómo determinar la probabilidad condicional se está en condiciones de abordar el cálculo de la probabilidad del *suceso producto*. Despejando $P(\boldsymbol{A}.\boldsymbol{B})$ de la expresión dada en la Definición 14, se obtiene

$$P(\boldsymbol{A}.\boldsymbol{B}) = P(\boldsymbol{A} / \boldsymbol{B}) \, . \, P(\boldsymbol{B})$$

Si ***A*** y ***B*** son sucesos independientes, por la Definición 12 resulta $P(\boldsymbol{A} / \boldsymbol{B}) = P(\boldsymbol{A})$, entonces

$$P(\boldsymbol{A}.\boldsymbol{B}) = P(\boldsymbol{A}).P(\boldsymbol{B})$$

Cuando los sucesos ***A*** y ***B*** son resultados de un experimento que consiste en la extracción de elementos de una población finita, para calcular la probabilidad del suceso producto (y la condicional) se debe tener en cuenta si el experimento se ha realizado *con reposición* o *sin reposición*.

Para comprender mejor este hecho se desarrollará el siguiente

Ejemplo 13: De un mazo de 52 cartas se seleccionan al azar dos cartas. ¿Cuál es la probabilidad de que ambas sean de diamante?.

Solución:

Para responder esta pregunta se distinguirán dos casos:

1. La primera carta extraída *es devuelta* al mazo antes de extraer la segunda.

2. La primera carta extraída *no es devuelta* al mazo y se extrae la segunda.

Sean los siguientes sucesos:

A = {obtener un diamante en la primera extracción}.
B = {obtener un diamante en la segunda extracción}.

La probabilidad buscada es $P(\boldsymbol{A}.\boldsymbol{B})$

1. $P(\boldsymbol{A})=13/52$ y $P(\boldsymbol{B} / \boldsymbol{A}) = P(\boldsymbol{B})=13/52$ porque la primera carta extraída es devuelta al mazo antes de extraer la segunda. Luego ***A*** y ***B*** son sucesos independientes y entonces

$$\begin{aligned} P(\text{obtener diamante en la primera y en la segunda extracción}) &= P(\boldsymbol{A}.\boldsymbol{B}) \\ &= P(\boldsymbol{A}).P(\boldsymbol{B}) \\ &= 169/2704 = 0.0625 \end{aligned}$$

2. $P(\boldsymbol{A})=13/52$ y $P(\boldsymbol{B} / \boldsymbol{A})=12/51$ porque la primera carta extraída no es devuelta al mazo y se extrae la segunda, lo que hace que ***A*** y ***B*** sean sucesos dependientes. Entonces

P(obtener diamante en la primera y en la segunda extracción) = P(***A.B***)
= P(***A***).P(***B/A***)
=156/2652 = 0.059

En general el experimento definido en el caso 1 se denomina *selección con reposición*, mientras que el caso 2 se denomina *selección sin reposición.*

3.7 Teorema de Bayes

El teorema de Bayes es un resultado muy utilizado para resolver problemas como el planteado a continuación

Ejemplo 14: En cierta investigación se desea calcular la probabilidad de que un animal presente una cierta enfermedad dado que está vacunado conociendo que la probabilidad de que esté vacunado dado que contrajo la enfermedad es 0.20.

Observar que se desea determinar una probabilidad condicional teniendo la probabilidad condicional inversa.

Para solucionar este problema es necesario enunciar el siguiente

Teorema: Sea ***E*** un experimento aleatorio, ***S*** el espacio muestral y sean $B_1, B_2,..., B_k$ sucesos mutuamente excluyentes (es decir $B_i \cap B_j = \emptyset$ para todo $i \neq j$) tales que ***S*** = $\bigcup_{i=1}^{k} B_i$. Entonces para cualquier suceso A del espacio muestral ***S***, se tiene que

$$P(B_j / A) = \frac{P(A / B_j).P(B_j)}{\sum_{i=1}^{k} P(A / B_i).P(B_i)}$$

Notar que $A = A \cap S = A \cap \left(\bigcup_{i=1}^{k} B_i \right) = \bigcup_{i=1}^{k} (A \cap B_i)$.

Así para solucionar el problema planteado vamos a utilizar el teorema enunciado (Mendenhall, W. et. al. 1994).

Solución:

Para este caso se tiene que el experimento aleatorio ***E*** es ”elegir al azar un animal”, luego al espacio muestral se lo puede pensar como la unión de los siguientes sucesos:
B_1={que el animal contraiga la enfermedad} y B_2={que el animal no contraiga la enfermedad}, entonces ***S*** = $B_1 \cup B_2$ y $B_1 \cap B_2 = \emptyset$. Sea el suceso A = {el animal está vacunado} entonces $A = A \cap$ ***S*** $= A \cap (B_1 \cup B_2) = (A \cap B_1) \cup (A \cap B_2)$ y aplicando probabilidad a ambos miembros se obtiene $P(A) = P(A/B_1).P(B_1) + P(A/B_2).P(B_2)$

Luego la expresión de la probabilidad que se desea determinar es

$$P(B_1 / A) = \frac{P(A / B_1).P(B_1)}{P(A / B_1).P(B_1) + P(A / B_2).P(B_2)}$$

En términos del problema se tiene:

La probabilidad de que un animal contraiga la enfermedad es 0.75 y la probabilidad de que un animal esté vacunado sabiendo que no contrajo la enfermedad es 0.70. Entonces

$$P(B_1 / A) = \frac{0.20 \, . \, 0.75}{0.20 \, . \, 0.75 + 0.70 \, . \, 0.25} = \frac{0.15}{0.325} \cong 0.46$$

Este resultado está indicando que *aproximadamente el 46 % de los animales que están vacunados presentan la enfermedad.*

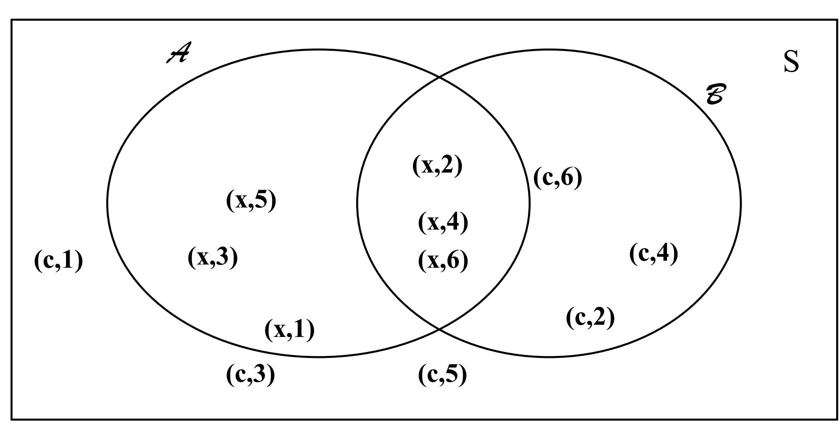
Ejercicios de Aplicación

1. *(Un poco de teoría de conjuntos y probabilidad)*

Sea el experimento de "*arrojar una moneda equilibrada y un dado regular en ese orden*".

a) En la figura siguiente se muestran el espacio muestral de este experimento y dos sucesos, A y B. Describir estos sucesos y calcular su probabilidad.

b) Calcular P(A. B) y P(A + B) e interpretar el resultado para este caso particular.



Espacio muestral para el Experimento "arroja una moneda y un dado".
(Con **x** y **c** se simbolizan "cruz" y "cara", respectivamente).

2.

Sean A y B dos características genéticas y supóngase que la probabilidad de que un animal manifieste la característica A es 1/2 , 3/4 de que manifieste la B y 3/8 que las manifieste a ambas. ¿ Cuál es la probabilidad de que un animal manifieste las características A o B ?.

3[2].

A continuación se da la clasificación de 872 campos seleccionados en una investigación sobre el rendimiento de arroz, clasificados de acuerdo al tipo de abono e irrigación.

[2] *En este problema y en los siguientes se supone que n es suficientemente grande como para verificarse la aproximación adecuada entre las frecuencias relativas y las probabilidades respectivas*.

IRRIGACIÓN / ABONO	Si	No
Sin Abono	123	413
Estiércol	81	223
Otros Abonos	14	18

a) Si se selecciona uno de esos campos al azar, hallar:

i. P("sin abono").
ii. P("de estar irrigado y con estiércol").
iii. P("sin abono o con estiércol, y no irrigado").
iv. P("de estar irrigado dado que no está abonado").
v. P("tener estiércol u otros abonos").

b) ¿Son independientes los sucesos "estar irrigado y estar abonado" ?.

4.

Un total de 86 bovinos afectados de fractura de primera falange fueron tratados con dos procedimiento específicos, uno tradicional y otro nuevo. Se confeccionó un registro para cada animal donde se dejaba constancia del tratamiento aplicado y de los resultados del mismo. Dichos registros arrojaron la siguiente clasificación: de los tratados con el procedimiento tradicional 42 se recuperaron y 9 no, en tanto que entre los tratados con el nuevo procedimiento 17 se recuperaron y 18 no.

Si se elige uno de los 86 registros al azar, encontrar la probabilidad de que el animal:

a) haya sido tratado con el nuevo procedimiento.

b) haya sido tratado con el procedimiento tradicional y se haya recuperado.

5.

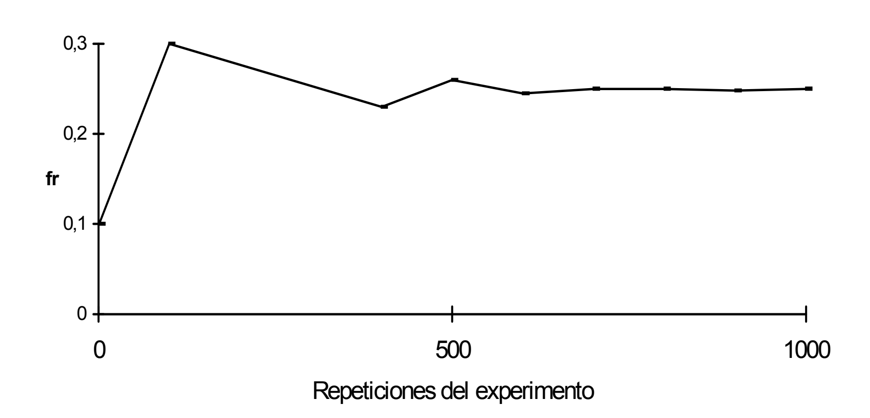
Considerar el experimento *aleatorio "registrar el sexo de las dos primeras crías de una vaca de cierta raza*" (suponga que $P(H)=P(M)= 1/2$).

a) Describir el espacio muestral (S) correspondiente a este experimento aleatorio.

b) Completar con las probabilidades requeridas la siguiente tabla:

Evento	Probabilidad
"que ambas crías sean machos"	
"que una cría sea hembra"	
"que sólo una cría sea macho"	

c) En el siguiente gráfico se informan los resultados correspondientes a una serie sucesiva de repeticiones del experimento y las correspondientes frecuencias relativas del evento "que ambas crías sean hembras". ¿Qué concepto teórico permite relacionar el comportamiento de las frecuencias relativas con el valor de la probabilidad del evento?.



Frecuencias relativas del evento “que ambas crías sean hembras” versus el número de repeticiones del experimento.

6.

Se cruzan dos organismos heterocigotos con genotipo Aa para cierta característica; un individuo de la primera generación de dicho cruzamiento puede poseer los siguientes cuatro genotipos:

	A	***a***
A	AA	Aa
a	aA	aa

a) Calcular la probabilidad de cada una de las combinaciones posibles, suponiendo que todas son igualmente probables.

b) En la descendencia anterior los individuos que son AA no pueden ser Aa o aa al mismo tiempo. ¿Cómo se denomina a los sucesos que mantienen esa relación y cómo se expresa matemáticamente?.

c) ¿Cuál es la probabilidad de que un individuo de la primera generación sea heterocigota?.

7.

En un heterocigota (Aa) al formarse los gametos puede ser que lleve el alelo "A" o el "a" con igual probabilidad. Otro heterocigota Aa formará las mismas gametas con igual probabilidad pero será un suceso independiente al anterior. Por lo tanto al producirse un cruzamiento la probabilidad de un genotipo cualquiera será igual al producto de las probabilidades genéticas que le darán origen.

a) Expresar la independencia mencionada en términos probabilísticos.

b) Hallar, utilizando la independencia, la probabilidad de un genotipo cualquiera para un individuo de la primera generación. Comparar con lo obtenido en el ejercicio anterior.

8.

Ciertos genes afectan el pelaje de los gatos domésticos. Uno de estos genes determina si el gato tienen pelaje blanco o blanco con manchas y otro de los genes determina si el pelaje es largo o corto. En la siguiente tabla se muestran algunas de las probabilidades de que un gato, miembro de la primera generación filial de un cruzamiento de gatos heterocigotas, posea alguna de las combinaciones según el color y tipo de pelaje:

TIPO DE PELAJE / COLOR DE PELAJE	Largo	Corto
Blanco		
Manchado	0.18	0.56

Completar con las probabilidades faltantes de tal manera que los eventos "que el gato posea un color de pelaje blanco" y "que el gato posea un pelaje largo" resulten independientes.

9.

La queratosis (anomalía de la piel) es debida a un gen dominante Q. Una mujer con queratosis cuyo padre era normal, se casa con un hombre con queratosis cuya madre era normal. Si esa pareja tiene 4 hijos. ¿Cuál es la probabilidad de que todos ellos tengan queratosis?.

10.

En el pollo las plumas sedosas están determinadas por un gen cuyo efecto es recesivo respecto al que rige las plumas normales. Si de un cruzamiento entre individuos heterocigotas se criasen 98 aves. Utilizando la definición estadística de probabilidad: ¿cuántos pollos cabría esperar que tuvieran plumas sedosas, y cuántos plumas normales?.

11.

En la producción de cerdos es estratégicamente importante el control de las enfermedades respiratorias, ya que éstas son causantes de pérdidas sustanciales. Una de tales enfermedades es la neumonía, y una manera de cuantificar su impacto es a través *del área pulmonar afectada*.
En un frigorífico se realizó, a través del tiempo, una clasificación de 10000 cerdos según la edad y el área pulmonar afectada por la neumonía, ambas convertidas en variables cualitativas. El siguiente cuadro describe los porcentajes de los animales clasificados según esas variables:

EDAD / ÁREA PULMONAR	Menor	Mayor
0-15 %	4000	4300
16% o más	300	1400

Suponiendo que los datos obtenidos en el frigorífico son una descripción acertada del estado de los cerdos de los criaderos de la zona, ¿cuál es la probabilidad de que un cerdo elegido al azar de dichos criaderos:

a) tenga un área pulmonar afectada del 16% o más?.

b) que sea de edad menor ?.

c) que sea de edad mayor y tenga un área pulmonar afectada de 15% o menos?.

d) ¿son independientes los eventos "que un cerdo posea 16% o más de área pulmonar afectada" y "que un cerdo sea de edad mayor".

4 Variables Aleatorias Discretas

Objetivos:

- Identificar variables aleatorias discretas.
- Construir la tabla de distribución de probabilidades de una variable aleatoria discreta.
- Reconocer los parámetros de las distribuciones de probabilidades.

4.1 Variable Aleatoria

Al describir el espacio muestral de un experimento, no se ha especificado que un resultado individual necesariamente deba ser un número. De hecho, se han citado varios ejemplos donde el resultado del experimento no fue una cantidad numérica. En esos casos la variable respuesta del experimento es una categoría, es decir que el espacio muestral correspondiente a ese experimento *no es un conjunto de números*. Sin embargo, en muchas situaciones experimentales se está interesado en asignar un número real a cada uno de los elementos del espacio muestral.

Para entender esta idea se presenta el siguiente

Ejemplo 1: Sea ***E***: "tirar un dado 2 veces" el experimento y ***S***={(1,1),(1,2),(1,3),...,(6,6)} su espacio muestral.

Se definen sobre este espacio muestral tres variables:

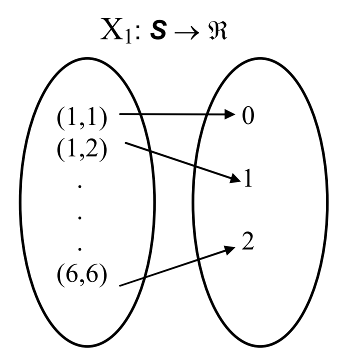
X_1: Cantidad de números pares que aparecen en la cara superior.

X_2: Diferencia entre los números que aparecen en la cara superior.

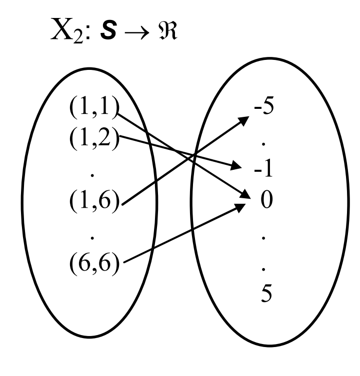
X_3: Producto de los números que aparecen en la cara superior.

Cada una de estas variables es una función que asigna a cada resultado del experimento aleatorio (elemento del espacio muestral) un número real.

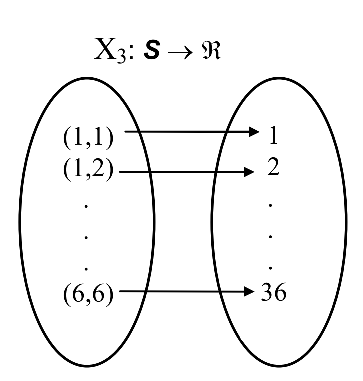
X_1: $\boldsymbol{S} \rightarrow \Re$



X_2: $\boldsymbol{S} \rightarrow \Re$



X_3: $\boldsymbol{S} \rightarrow \Re$



A partir de esta idea intuitiva se puede dar la siguiente

Definición 1: Sea ***E*** un experimento aleatorio y ***S*** el espacio muestral asociado con dicho experimento. Una función X que asigna a cada uno de los elementos s de ***S*** un número real $X(s)$ se llama *variable aleatoria*, lo que se puede expresar en símbolos

$$X: \boldsymbol{S} \rightarrow \Re$$
$$s \mapsto X(s)$$

El conjunto de valores posibles que puede asumir una variable aleatoria es llamado recorrido o imagen de la variable, denotado por $R(X)$. Para el Ejemplo 1

$R(X_1) = 0, 1, 2$
$R(X_2) = -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$
$R(X_3) = 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30, 36$

A cada valor de la variable se le puede asociar su probabilidad de ocurrencia y en base a ellos se construye una tabla, llamada *tabla de distribución de probabilidades.*

Para el Ejemplo 1 las tablas correspondientes a las variables X_1 y X_2 son:

X_{1i}	$p_i=P(X_1=x_{1i})$
0	9/36
1	18/36
2	9/36
	1

x_{2i}	$p_i=P(X_2=x_{2i})$
-5	1/36
-4	2/36
-3	3/36
-2	4/36
-1	5/36
0	6/36
1	5/36
2	4/36
3	3/36
4	2/36
5	1/36
	1

Para ejemplificar el cálculo de las probabilidades, puede considerarse el caso $P(X_1=2)$. Dado que el valor $X_1=2$ está asociado al suceso {(2,2), (2,4), (2,6), (4,2), (4,4), (4,6), (6,2), (6,4), (6,6)} resulta $P(X_1=2)=9/36$.

En cuanto a la notación, se utilizarán mayúsculas, como X, para denotar variables aleatorias y minúsculas, como x_i, para representar los valores particulares que puede tomar una variable aleatoria. Por su parte p_i representa la $P(X=x_i)$.

Notar que en cada tabla de distribución de probabilidades, *la suma de las probabilidades* p_i es igual a 1.

4.2 Variable Aleatoria Discreta

Definición 2: Una variable aleatoria X se llama *discreta* si solamente puede tomar un número *finito o infinito numerable* de valores distintos.

Las variables X_1, X_2 y X_3 del Ejemplo 1 son variables aleatorias discretas.

Definición 3: Sea X una variable aleatoria discreta, se llama *función de probabilidad puntual* de la variable X a la función P, que asigna a cada resultado posible x_i un número $p_i=P(X=x_i)$, el cual verifica las siguientes condiciones

1. $0 \leq p_i \leq 1$ para todo i.

2. $\sum_{i=1}^{k} p_i = 1$ donde k indica el número de valores diferentes que toma la variable.

En general a la colección de pares (x_i, p_i) con $i=1, 2, 3,..., k$ se llama *Distribución de Probabilidad Puntual* .

Es importante observar la diferencia entre la tabla de distribución de probabilidades y la tabla de distribución de frecuencias. La presentada aquí contiene *todos* los valores posibles de la variable con sus respectivas probabilidades, mientras que en la tabla de distribución de frecuencias aparecen *sólo* los valores de variable que se dieron en la muestra y sus frecuencias absolutas y relativas. Los valores $\mathbf{f_{ri}}$ y $\mathbf{p_i}$ son proporciones, la diferencia está en que la primera es con respecto al tamaño de la muestra mientras que la segunda es con respecto a toda la población. Además, para construir la tabla de distribución de probabilidades *no* es necesario realizar el experimento mientras que para construir la tabla de distribución de frecuencias *sí* lo es.

La información dada por la tabla de distribución de probabilidades se puede representar usando gráficos, como los que corresponden a una tabla de frecuencias no agrupadas. Por ejemplo para la variable X_1 el diagrama de barras es el que se muestra a continuación. En el eje de las abscisas se representan los distintos valores de la variable X_1 y en el eje de las ordenadas las probabilidades con que aparecen dichos valores.

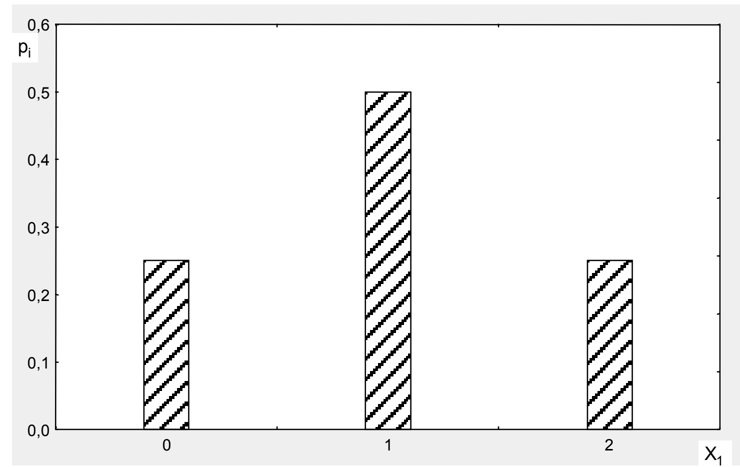


Gráfico 1: Distribución de probabilidades de la cantidad de números pares

4.3 Características numéricas de la variable aleatoria

Así como se puede resumir la información de una *muestra estadística* mediante números o valores típicos llamados *estadísticos*, también se puede resumir la información de una *población estadística* mediante números llamados *características numéricas* de la variable.

Las características numéricas de una población estadística o valores típicos de una población más utilizados son la Esperanza o Media, la Varianza y la Desviación Estándar. Estos valores se obtienen a partir de fórmulas análogas a las que permiten obtener los estadísticos utilizando la información de toda la población. La definición de cada una de estas características está dada a continuación.

Definición 4: Si X es una variable aleatoria discreta, entonces

1. La esperanza o media poblacional está dada por $E(X) = \sum_{i=1}^{\infty} p_i\, X_i$
2. La varianza poblacional está dada por $Var(X) = \sum_{i=1}^{\infty} p_i\, (X_i - E(X))^2$
3. $Des(X) = +\sqrt{Var(X)}$

Es fácil observar la similitud de estas características numéricas con los estadísticos $\overline{X}$, S^2 y S. A continuación, se muestra la similitud existente entre $\overline{X}$ y $E(X)$ para lo cual es importante recordar que $f_{ri} \cong p_i$, cuando **n** tiende a infinito.

$$\overline{X} = \sum_{i=1}^{k} \frac{f_i X_i}{n} = \sum_{i=1}^{k} \frac{f_i}{n} X_i = \sum_{i=1}^{k} f_{ri} X_i \cong \sum_{i=1}^{k} p_i X_i = E(X)$$, cuando **n** es grande.

4.3.1 Propiedades de la Esperanza y la Varianza

Si X e Y son variables aleatorias y **c** es una constante, entonces

E_1. $E(c) = c$

E_2. $E(c.X) = c.E(X)$

E_3. $E(X \pm Y) = E(X) \pm E(Y)$

E_4. Si X e Y además son variables aleatorias independientes, entonces $E(X.Y)=E(X).E(Y)$

V_1. $Var(c) = 0$

V_2. $Var(c.X) = c^2.Var(X)$

V_3. Si X e Y además son variables aleatorias independientes, entonces $Var(X \pm Y)=Var(X) + Var(Y)$

Ejemplo 2: Calcular la $E(X_1)$ y $Var(X_1)$ para la variable X_1 definida anteriormente.

$$E(X_1) = \frac{9}{36} \bullet 0 + \frac{18}{36} \bullet 1 + \frac{9}{36} \bullet 2 = \frac{36}{36} = 1$$

Que la $E(X_1) = 1$ significa que si *se repite muchas veces el experimento de arrojar dos veces un dado y se cuentan en cada repetición los números pares que aparecen en la cara superior, entonces la cantidad de números pares en cada repetición es aproximadamente* 1.

$$\mathrm{Var}(X_1) = \frac{9}{36} \bullet (0-1)^2 + \frac{18}{36} \bullet (1-1)^2 + \frac{9}{36} \bullet (2-1)^2 = \frac{9}{36} \bullet 1 + \frac{18}{36} \bullet 0 + \frac{9}{36} \bullet 1 = 0.5$$

$$\mathrm{Des}(X_1) = +\sqrt{0.5} \cong 0.71$$

Muchos son los experimentos que se pueden ajustar a distintos modelos probabilísticos para variables aleatorias de tipo discreto, entre ellos se estudiarán en detalle los Modelos Probabilísticos *Bernoulli*, *Binomial*, *Hipergeométrico*, *Poisson*, *Geométrico* y *Binomial Negativo.*

4.4 Modelo Probabilístico Bernoulli

En primer lugar se desarrollará el modelo básico para generar los modelos mencionados anteriormente.

Ejemplo 3: Un vendedor de semillas de cierta especie garantiza un poder germinativo del 90%. Una persona selecciona 1 semilla y la siembra en una maceta. ¿Cuál es la probabilidad de que germine ?.

Solución: En este caso

E: "sembrar 1 semilla" ***S*** = {G,NG}

X: número de semillas que germinan R(X) = 0,1

$$P(\text{que una semilla germine}) = P(X=1) = 0.90$$

La distribución de probabilidades de la variable X es:

Tabla 1: Distribución de probabilidades del Número de semillas que germinan

X_i	Elementos de ***S***	$p_i = P(X=x_i)$
0	NG	0.10
1	G	0.90

Definición 5: Se dice que un experimento es un Ensayo de Bernoulli si tiene sólo dos resultados posibles, uno se denominará éxito (E) y el otro fracaso (F).

En este caso el espacio muestral es ***S*** = {E,F}.

Llamando p=P(Exito) y q=P(Fracaso) y teniendo en cuenta que P(Exito)+P(Fracaso)=1 (o sea p+q=1) se obtiene que q=1–p.

En el Ejemplo 3 resulta q=P(Fracaso)=P(una semilla no germine)=1–0.90=0.10.

Definición 6: El *parámetro* del modelo Bernoulli es p (probabilidad del suceso éxito).

Para el Ejemplo 3, el suceso éxito es "la semilla germina" y el parámetro es p=0.90.

Definición 7: La *variable aleatoria Bernoulli* X_{BER} cuenta el número de éxitos cuando se realiza un Ensayo de Bernoulli.

El *conjunto de valores* que asume es $R(X_{BER})=0,1$.

Definición 8: La probabilidad de obtener **k** éxitos en un Ensayo de Bernoulli está dada por

$$P(X_{BER}=k) = p^k (1-p)^{1-k} \text{ con } k=0,1$$

Notar que $\sum_{k=0}^{1} p^k (1-p)^{1-k} = 1$.

4.4.1.Características Numéricas

Las *características numéricas* de la variable aleatoria Bernoulli son:

1. $E(X_{BER})=p$
2. $Var(X_{BER})=p \bullet q$
3. $Des(X_{BER}) = \sqrt{p \bullet q}$

El resultado de cada una de ellas es inmediato.

Algunos experimentos no responden en forma inmediata a las características de un Ensayo de Bernoulli. Por ejemplo el experimento ***E***: "tirar un dado equilibrado" tiene más de dos resultados posibles, o sea ***S***={1,2,3,4,5,6}. Pero si se definen E: {sale un número cinco} y F: {sale un número distinto de cinco} puede pensarse al experimento con sólo dos resultados posibles, verificando así la condición deseada. Por tanto es posible transformar algunos experimentos en Ensayos de Bernoulli de forma conveniente, de acuerdo a lo que se desea determinar.

4.5 Modelo Probabilístico Binomial

El modelo binomial se ajusta a muchas situaciones prácticas. Para presentarlo será utilizado el siguiente

Ejemplo 4: Un vendedor de semillas de cierta especie garantiza un poder germinativo del 90%. Una persona compra un paquete de 4 semillas y las siembra en una maceta. ¿Cuál es la probabilidad de que germinen 3, y de que germinen por lo menos 2?

Solución: En este caso

E: "sembrar 4 semillas"

Este experimento se puede pensar como 4 repeticiones del Ensayo de Bernoulli definido en el Ejemplo 3, luego el espacio muestral asociado a él es:

S={ (G,G,G,G), (G,G,G,NG), (G,G,NG,G), (G,NG,G,G), (NG,G,G,G), (G,G,NG,NG), (G,NG,NG,G), (NG,NG,G,G), (NG,G,G,NG), (NG,G,NG,G), (G,NG,G,NG), (G,NG,NG,NG), (NG,G,NG,NG), (NG,NG,G,NG), (NG,NG,NG,G), (NG,NG,NG,NG) }

Considerando que el suceso éxito es "la semilla germina" se define la variable

X: "Número de semillas que germinan de un paquete de 4" cuyo recorrido es $R(X)=0,1,2,3,4$.

La probabilidad de que germinen 3 semillas es:

$$\begin{aligned} P(X=3) &= P((G,G,G,NG)) + P((G,G,NG,G)) + P((G,NG,G,G)) + P((NG,G,G,G)) \\ &= P(G).\ P(G).\ P(G).\ P(NG) + ... + P(NG).\ P(G).\ P(G).\ P(G) \\ &= 0.9 \bullet 0.9 \bullet 0.9 \bullet 0.1 + ... + 0.1 \bullet 0.9 \bullet 0.9 \bullet 0.9 \\ &= 4 \bullet 0.9^3 \bullet 0.1 \\ &= 0.2916 \end{aligned}$$

Este resultado es válido bajo el supuesto de que la germinación de una semilla en una repetición cualquiera es independiente de la germinación de otra semilla.

La probabilidad encontrada significa que s*i se sembraran muchas veces 4 semillas, aproximadamente en el 29 % de los casos 3 semillas germinan.*

En la siguiente tabla se presenta la distribución de probabilidades de la variable X: Número de semillas que germinan de un paquete de 4 semillas.

Tabla 2: Distribución de probabilidades del Número de semillas que germinan

X_i	Elementos de ***S***	$p_i = P(X=x_i)$
0	(NG,NG,NG,NG)	$P(X = 0) = 0.1^4 = 0.0001$
1	(G,NG,NG,NG) (NG,G,NG,NG) (NG,NG,G,NG) (NG,NG,NG,G)	$P(X = 1) = 4 \bullet 0.9 \bullet 0.1^3 = 0.0036$
2	(G,G,NG,NG) (G,NG,NG,G) (NG,NG,G,G) (NG,G,G,NG) (NG,G,NG,G) (G,NG,G,NG)	$P(X = 2) = 6 \bullet 0.9^2 \bullet 0.1^2 = 0.0486$
3	(G,G,G,NG) (G,G,NG,G) (G,NG,G,G) (NG,G,G,G)	$P(X = 3) = 4 \bullet 0.9^3 \bullet 0.1 = 0.2916$
4	(G,G,G,G)	$P(X = 4) = 0.9^4 = 0.6561$

En base a los datos de la tabla la probabilidad de que por lo menos 2 semillas germinen es:

$$P(X \geq 2) = P(X=2) + P(X=3) + P(X=4) = 0.0486 + 0.2916 + 0.6561 = 0.9963$$

Esta probabilidad indica que s*i se sembraran muchas veces 4 semillas, aproximadamente en el 99% de los casos hay 2 o más semillas que germinan.*

Los valores de la variable X y sus respectivas probabilidades pueden ser representados en un diagrama de barras tal como se hizo para la variable X_1.

Las características numéricas en este caso son:

1. $E(X) = \sum_{i=1}^{5} x_i p_i = 3.6$, valor que se interpreta de la siguiente manera: s*i se sembraran muchas veces 4 semillas, aproximadamente germinan entre* 3 *y* 4 *semillas.*

La expresión "si se sembraran 4 semillas muchas veces" significa que el experimento Binomial "sembrar 4 semillas" se repite muchas veces.

2. $Var(X) = \sum_{i=1}^{5} (x_i - 3.6)^2 p_i = 0.36$

3. $Des(X)=0.6$.

El experimento definido en el Ejemplo 4 tiene las siguientes características:

- El experimento E_i :“elegir una semilla y sembrarla” tiene dos resultados posibles: la semilla germina o la semilla no germina (Ensayo de Bernoulli).
- El experimento E_i se realiza un número fijo de veces, 4.
- Los resultados de los ensayos E_i son independiente entre sí, pues el hecho que una semilla germine o no germine no implica que otra germine o no germine.
- La probabilidad de que una semilla germine es 0.90 (el poder germinativo de la semilla es del 90%) se mantiene igual cada vez que se siembra una semilla.

Experimentos como éste (elegir 4 semillas y sembrarlas) que cumplen con las características mencionadas, se ajustan al *Modelo Binomial*.

Definición 9: Se dice que un *experimento aleatorio se ajusta al Modelo Binomial* si posee las siguientes características:

1. Consta de **n** repeticiones del Ensayo de Bernoulli.
2. Los resultados de las repeticiones del Ensayo de Bernoulli son *independientes* entre sí (el resultado de una repetición no influye en el de ninguna otra).
3. La probabilidad del éxito **p** en cada repetición del ensayo permanece *constante* (no cambia de una repetición a otra).

Cuando el experimento consiste en realizar extracciones de una población pequeña, para que se cumplan las características 2 y 3, dichas extracciones deben ser realizadas con reposición.

Definición 10: Los *parámetros* del Modelo Binomial son **n** (número de repeticiones del Ensayo de Bernoulli) y **p** (probabilidad de éxito).

Para el Ejemplo 4 los parámetros son $n=4$ y $p=0.90$.

Cada valor de **n** y **p** identifican un Modelo Binomial particular (por ser los parámetros de la distribución), por lo tanto existen infinitos Modelos Binomiales dado que **n** puede tomar cualquier valor en el conjunto de los números naturales y **p** cualquier número real entre 0 y 1.

Definición 11: La *variable aleatoria binomial* X_b cuenta el "número de éxitos en las **n** repeticiones del Ensayo de Bernoulli".

El conjunto de valores que asume la variable X_b es $R(X_b) = 0,1,2,3,...,n$.

Que la variable X_b tome el valor 0 significa que no hubo éxitos en las **n** repeticiones del Ensayo de Bernoulli y que tome el valor **n** significa que hubo **n** éxitos en las **n** repeticiones del ensayo.

Para el Ejemplo 4 la variable aleatoria X_b: "Número de semillas que germinan entre las 4 semillas sembradas" es una variable aleatoria binomial y $R(X_b) = 0,1,2,3,4$.

Para determinar la probabilidad de que la variable X_b asuma el valor **k**, se debe en primer lugar dar la siguiente

Definición 12: El *número combinatorio* $\binom{n}{k}$ es la cantidad de subconjuntos con **k** elementos que pueden obtenerse de un conjunto de **n** elementos. En símbolos

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

donde $n! = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot 2 \cdot 1$ es el *factorial* del número natural n (número que se obtiene al realizar el producto de los primeros n números naturales). Por convención se define $0! = 1$.

Se define a continuación la forma de calcular la probabilidad buscada.

Definición 13: La probabilidad de obtener **k** éxitos en las **n** repeticiones del Ensayo de Bernoulli es:

$$P(X_b = k) = \binom{n}{k} p^k \, q^{n\text{-}k} \qquad \text{con } k = 0,1,2,...,n \qquad (4.1)$$

De acuerdo a la Definición 3, $\sum_{k=0}^{n} P(X_b = k) = \sum_{k=0}^{n} \binom{n}{k} p^k \, q^{n\text{-}k} = 1$

Para el Ejemplo 4, usando la Definición 13, la probabilidad de que germinen 2 de las 4 semillas es:

$$P(X_b = 2) = \binom{4}{2} 0.90^2 \; 0.10^{4\text{-}2} = 0.0486$$

donde el número combinatorio que aparece en la fórmula anterior es:

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = 6$$

Este resultado coincide con el número de sucesos elementales que contienen dos éxitos (G) y dos fracasos (NG) en la Tabla 2. Por su parte la probabilidad coincide con la presentada en dicha tabla.

Ejemplo 5: En la situación del ejemplo anterior se podría haber definido la siguiente variable aleatoria binomial

X_b^* : "Número de semillas que no germinan entre las 4 sembradas".

o sea que en este caso el suceso éxito es "la semilla no germina" y por lo tanto

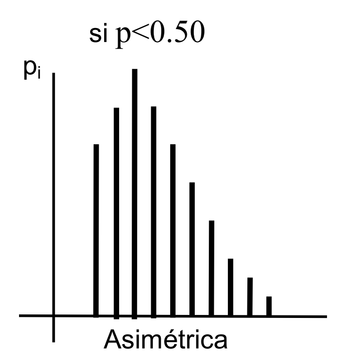
p=P(Éxito)=P(La semilla no germina)=0.10.

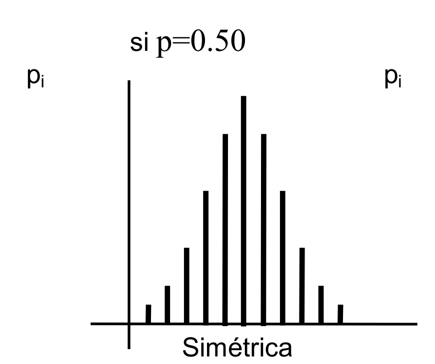
Los parámetros del modelo probabilístico asociado a la variable X_b^* son $n=4$ y $p=0.10$, y los valores de la variable son $0,1,2,3,4$.

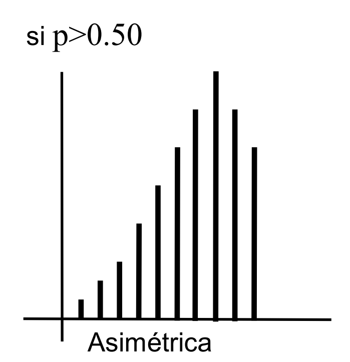
Ahora debe notarse que en un paquete de 4 semillas, el suceso "1 semilla no germina" es equivalente al suceso "3 semillas germinan". Por lo tanto la probabilidad de estos dos sucesos es la misma, es decir $P(X_b^* = 1) = P(X_b = 3) = 0.2916$.

Esto lleva a concluir que la probabilidad no varía al cambiar el éxito, usando correctamente los elementos que intervienen para su cálculo.

Para un valor fijo de **n** se tiene la distribución de probabilidades de una variable aleatoria binomial, como se muestra en los siguientes diagramas de barras.







4.5.1. Características numéricas

A continuación se muestran las *características numéricas* de la variable aleatoria binomial. Para cada una de ellas, el resultado final es presentado sin demostración, ya que la misma requiere de algunos conceptos que este texto no contempla (Meyer, P. 1992). Luego:

1. $E(X_b) = \sum_{k=0}^{n} k\,P(X_b = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = n \bullet p$
2. $Var(X_b) = n \bullet p \bullet q$
3. $Des(X_b) = \sqrt{n \bullet p \bullet q}$

4.5.2. Uso de Tabla

Para calcular las probabilidades de que una variable aleatoria binomial tome ciertos valores se dispone de una tabla para distintos valores de **n** y **p**.

La Tabla A del Apéndice corresponde a la distribución binomial. En la primera fila se muestran los valores de p=P(Éxito), en la primera columna los valores de **n** y en la segunda columna los valores de la variable X correspondientes a cada valor de **n**. Para valores determinados de **n** y **p**, la probabilidad de que ocurran **k** éxitos está dada por el número que se encuentra en la intersección de la fila y columna correspondientes.

4.6 Modelo Probabílistico Hipergeométrico

Este modelo describe experimentos que consisten en una secuencia de extracciones (Ensayos de Bernoulli) de una población finita, sin reposición. Una aplicación de este modelo se presenta en el siguiente

Ejemplo 6: En un vivero hay 100 plantines de una misma especie, de los cuales se sabe que 40 están atacados. Se extraen 3 plantines al azar, sin reposición. Se desea calcular la probabilidad de obtener dos plantines atacados entre los 3 extraídos.

Solución:

El experimento ***E***: "Extraer 3 plantines sin reposición".
La variable aleatoria X: Número de plantines atacados entre los 3 extraídos.
$R(X)=0, 1, 2, 3.$

Si se observan las características de este experimento, se ve que tiene mucho en común con un experimento que se ajuste al Modelo Binomial; lo único que lo diferencia de aquel es que aquí las $\mathbf{n}$ repeticiones del Ensayo de Bernoulli (extracciones) son sin reposición, luego la probabilidad de éxito no es constante y las repeticiones no son independientes.

Cuando un experimento consiste en extraer sin reposición $\mathbf{n}$ elementos de una población finita de $\mathbf{N}$ elementos de los cuales $\mathbf{r}$ son éxitos, se ajusta al *Modelo Probabilístico Hipergeométrico.*

Definición 14: Los *parámetros* de esta distribución son $\mathbf{N}$ (número total de elementos de la población), $\mathbf{n}$ (número de extracciones o ensayos a realizar) y $\mathbf{r}$ (número de éxitos en la población).

Definición 15: Una *variable aleatoria hipergeométrica* X_h cuenta el número de éxitos en $\mathbf{n}$ extracciones sin reposición (Ensayos de Bernoulli).

El conjunto de valores que asume X_h es $R(X_h)=0, 1, 2, 3,..., n.$

Definición 16: Sea X_h una variable aleatoria con distribución hipergeométrica. La probabilidad de obtener $\mathbf{k}$ éxitos en las $\mathbf{n}$ extracciones sin reposición es

$$P(X_h = k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}} \qquad (4.2)$$

con $k=0, 1, 2, 3,..., n$ (si $n \leq r$) o $k=0, 1, 2, 3,..., r$ (si $n > r$) porque no pueden lograrse más éxitos de los que hay en todo el conjunto.

De acuerdo a la Definición 3, $\sum P(X_h = k) = \sum \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{n}{N}} = 1$

Para el Ejemplo 6 los parámetros son $N=100$, $r=40$, $n=3$ y X_h: Número de plantines atacados entre los 3 extraídos. Entonces la probabilidad de que 2 plantines estén atacados es

$$P(X_h = 2) = \frac{\binom{40}{2}\binom{60}{1}}{\binom{100}{3}} = 0.2894$$

Este valor indica que si *se extrajeran 3 plantines muchas veces de una población de 100 plantines de los cuales 40 están afectados, en aproximadamente el 29 % de los casos 2 están atacados.*

4.6.1 Características Numéricas

Definiendo $p = \frac{r}{N}$ como la proporción de éxitos en la población y $q=1-p$ como la proporción de fracasos en la población, se definen

1. $E(X_h) = n \bullet p$
2. $Var(X_h) = \frac{N-n}{N-1} \bullet n \bullet p \bullet q$
3. $Des(X_h) = \sqrt{\frac{N-n}{N-1} \bullet n \bullet p \bullet q}$

Para llegar a estos resultados basta reemplazar en la definición de las características numéricas de una variable aleatoria discreta la $P(X_h=k)$ cuya expresión está dada en (4.2).

4.6.2 Relación entre Modelo Hipergeométrico y Modelo Binomial

a) Cuando el tamaño de la muestra es pequeño con respecto al tamaño de la población, la probabilidad de obtener éxito permanecerá aproximadamente igual prueba a prueba, sin importar los resultados de las pruebas anteriores.
Por ejemplo si se tienen 10000 plantines de los cuales 6000 están atacados y se extraen **n** plantines sin reposición ($N=10000$, $r=6000$):

$P($primer plantín esté atacado$) = \frac{r}{N} = \frac{6000}{10000} = 0.6$

$P($segundo plantín esté atacado/primer plantín está atacado$) = \frac{r-1}{N-1} = \frac{5999}{9999} = 0.59999 \cong 0.6$

$P($segundo plantín esté atacado/primer plantín no está atacado$) = \frac{r}{N-1} = \frac{6000}{9999} = 0.60006 \cong 0.6$

$P($tercer plantín esté atacado/primero y segundo atacados$) = \frac{5998}{9998} = 0.5999 \cong 0.6$

$P($tercer plantín esté atacado/primero atacado y segundo no atacado o primero no atacado y segundo atacado$) = \frac{5999}{9998} = 0.60002 \cong 0.6$

$P($tercer plantín esté atacado/primero y segundo no atacados$) = \frac{6000}{9998} = 0.60001 \cong 0.6$

Luego en este caso se puede asumir que la variable en estudio se ajusta al Modelo Probabilístico Binomial.

b) Cuando el tamaño de muestra es relativamente grande con respecto al tamaño de la población, la probabilidad de éxito en la primera extracción $\frac{r}{N}$ diferirá considerablemente de la probabilidad de éxito en la segunda extracción $\frac{r-1}{N-1}$ o $\frac{r}{N-1}$. Por ejemplo para $N=10$ y $r=6$, la probabilidad de que el primer plantín esté atacado es $\frac{6}{10} = 0.6$, en tanto que la probabilidad

de que el segundo plantín esté atacado dado que el primero está atacado es $\frac{5}{9} = 0.55$ mientras que la probabilidad de que el segundo plantín esté atacado dado que el primero no está atacado es $\frac{6}{9} = 0.66$. Para la tercera extracción las probabilidades son 0.5, 0.625 y 0.75, respectivamente.

En este caso la probabilidad de éxito difiere considerablemente de un ensayo a otro y por lo tanto no se las puede suponer iguales en cada repetición, por lo que el Modelo Probabilístico Hipergeométrico parece el adecuado.

4.7 Modelo Probabilístico de Poisson

Otro modelo muy utilizado para variables aleatorias discretas es el Modelo de Poisson. Para describirlo se tomará como base el siguiente

Ejemplo 7: La deficiencia en el número de glóbulos rojos en la sangre puede determinarse mediante el examen microscópico de una muestra de sangre. Suponiendo que un pequeño volumen de sangre contiene en promedio 10 glóbulos rojos para personas normales, ¿cuál es la probabilidad de que una muestra de sangre de una persona normal contenga 7? ¿y de que contenga 7 glóbulos rojos o menos?.

Solución:

El experimento aleatorio ***E***: "Sacar una muestra de sangre".
La variable aleatoria X: Número de glóbulos rojos en una muestra de sangre.
$R(X)=0, 1, 2, 3, ...$

Observar las siguientes características de este experimento:

1. No se puede fijar el número máximo de glóbulos rojos que pueden aparecer en la muestra. Lo que sí se sabe es el *número promedio de glóbulos rojos para una muestra de sangre de una persona normal*; en este caso es 10 y se lo denomina λ.

2. Se puede determinar el número de glóbulos rojos que hay (número de éxitos) en la muestra de sangre, pero no el número de glóbulos rojos que no hay (número de fracasos) en la muestra de sangre.

Las características 1 y 2 indican claramente que la variable X no sigue el modelo binomial, por lo tanto para calcular las probabilidades pedidas se necesita buscar un modelo probabilístico adecuado para esta situación.

4.7.1 Supuestos del Modelo Poisson

Algunas de las características (o condiciones) bajo las cuales se puede esperar que un experimento pueda ser descripto por el modelo de Poisson son:

1. Los sucesos que ocurren en un intervalo de tiempo (o región del espacio o volumen) son independientes de los que ocurren en cualquier otro intervalo de tiempo (área o volumen) independientemente de como se elige el intervalo.

2. La probabilidad de que un suceso se presente es proporcional a la longitud del intervalo de tiempo (volumen o región del espacio).
3. La probabilidad de que dos o más sucesos se presenten en un intervalo de tiempo muy pequeño (área o volumen) es despreciable, por esta razón es que suele llamarse ley de los sucesos raros.

Sobre experimentos que verifiquen estas condiciones pueden definirse variables aleatorias como las siguientes:

* Número de bacterias en un cultivo dado.
* Número de plantas de musgo sobre área determinada de una colina.
* Número de parásitos que habitan en un huésped.
* Número de casos de gripe manifestados en una ciudad durante una semana.
* Número de mutaciones ocurridas en una cadena genética en el intervalo de un mes.

Para todos los ejemplos anteriores la variable cuenta el número de éxitos en intervalo de tiempo, área o volumen.

Definición 17: El *parámetro* del modelo de Poisson es λ que indica el número promedio de éxitos en cada intervalo de tiempo, área o volumen.

Definición 18: La *variable aleatoria de Poisson* X_p cuenta el número de éxitos en un intervalo de tiempo, área o volumen.

El conjunto de valores que asume X_p es $R(X_p)=0, 1, 2, 3,...$

Definición 19: Sea X_p una variable aleatoria con distribución de Poisson. La probabilidad de obtener $\mathbf{k}$ éxitos en un intervalo de tiempo, área o volumen es

$$P(X_p = k) = \frac{e^{-\lambda} \bullet \lambda^k}{k!} \quad \text{con } k=0, 1, 2, 3,... \tag{4.3}$$

De acuerdo a la Definición 3 $\sum_{k=0}^{\infty} P(X_p = k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda} \bullet \lambda^k}{k!} = 1$

Continuando el Ejemplo 7 resulta $P(X_P = 7) = \frac{10^7 \bullet e^{-10}}{7!} = 0.09008$

Este valor indica que si *se extrajeran muchas muestras de sangre aproximadamente el 9 % de ellas tienen 7 glóbulos rojos.*

$$P(X_P \le 7) = \sum_{k=0}^{7} \frac{10^k \bullet e^{-10}}{k!} = 0.22022$$

Este valor indica que si *se extrajeran muchas muestras de sangre aproximadamente el 22 % de ellas tienen 7 glóbulos rojos o menos.*

4.7.2 Características Numéricas

Las características numéricas de una variable aleatoria discreta Poisson son:

1. $E(X_p)=\lambda$
2. $Var(X_p)= \lambda$
3. $Des(X_p)=\sqrt{\lambda}$

Para llegar a estas igualdades basta reemplazar en la definición de las características numéricas de una variable aleatoria discreta la expresión (4.3).

En la variable Poisson la esperanza y la varianza coinciden con el valor del parámetro de la distribución.

4.7.3 Relación entre Modelo Binomial y Modelo Poisson

En las aplicaciones típicas del modelo Binomial el número de veces que se repite el Ensayo de Bernoulli (**n**) es relativamente pequeño, pero es frecuente que se presenten situaciones en que se aplica la distribución binomial para un valor de **n** *grande* y **p** muy pequeño. En este caso el cálculo de probabilidad para un valor de variable dado es difícil de determinar manualmente, por lo que se encontró una forma aproximada de resolverla usando el modelo de Poisson (Meyer, P.1992). Se puede demostrar que: $P(X_b=k) \cong P(X_p=k)$, o sea

$$P(X_b = k) = \binom{n}{k} p^k q^{n-k} \cong \frac{\lambda^k e^{-\lambda}}{k!}, \qquad \text{con } \lambda = n \bullet p$$

Ejemplo 8: Si en la fabricación de piezas bajo control se sabe que la probabilidad de que aparezca una pieza defectuosa es $p=0.01$ y se reúnen las piezas en cajas de $n=200$ unidades, para calcular la probabilidad de obtener 10 piezas defectuosas es conveniente usar una aproximación debida a Poisson

Entonces

$$P(X_b = 10) = \binom{200}{10} 0.01^{10} 0.99^{190} = \frac{2^{10} e^{-2}}{10!} = 0.000037, \qquad \text{con } \lambda = 200 \bullet 0.01 = 2$$

4.7.4 Uso de tabla

En el Apéndice se encuentra la Tabla B correspondiente a la distribución de Poisson. En la primera columna de la tabla se muestran diferentes valores de λ y en la primera fila distintos valores de X. Luego para un valor determinado de λ, la probabilidad de que ocurran **k** éxitos está dada por el número que se encuentra en la intersección de la fila y columna correspondiente.

4.8 Modelo Probabilístico Geométrico

El Modelo Probabilístico Geométrico describe experimentos que son una secuencia de Ensayos de Bernoulli independientes con parámetro **p** hasta obtener 1 éxito. Para estudiar este modelo se presenta el siguiente

Ejemplo 9: Se sabe que en una parcela el 20% de las plantas están atacadas, y se desea calcular la probabilidad de obtener la primera planta atacada en la tercera extracción.

Solución:
El experimento es ***E***: “extraer sucesivamente una planta de una parcela hasta obtener una planta atacada”.
La variable aleatoria X: Número de extracciones necesarias hasta que salga una planta atacada.
$R(X)=1, 2, 3, ...$

El experimento no responde a ninguno de los descriptos anteriormente, aunque tiene algunas características en común con algunos de ellos. Para calcular las probabilidades sobre variables como ésta, se describe un nuevo modelo probabilístico.

Cuando un experimento consiste en repeticiones independientes de un Ensayo de Bernoulli hasta que ocurra el primer éxito, se dice que se ajusta a un *Modelo Geométrico*.

Definición 20: El *parámetro* del Modelo Geométrico es $\mathbf{p}$ (probabilidad de éxito).

Definición 21: La *variable aleatoria geométrica* X_g cuenta el número de ensayos requeridos hasta que sale el suceso éxito por primera vez.

El conjunto de valores que asume es $R(X_g)=1, 2, 3,....$

Definición 22: Sea X_g una variable aleatoria con distribución geométrica. La probabilidad de que se realicen $\mathbf{k}$ ensayos hasta obtener el suceso éxito por primera vez es

$$P(X_g=k) = q^{k-1} \cdot p \qquad \text{con } k=1, 2, 3,... \tag{4.4}$$

Donde $\mathbf{q}$ es la probabilidad de fracaso en cada repetición (igual que en la distribución binomial).

Retomando el Ejemplo 9 la probabilidad de obtener la primera planta atacada en la tercera extracción es

$$P(X_g=3) =(1–0.20)^{3-1} \cdot 0.20 = 0.128$$

Este valor indica que *si se realiza muchas veces el experimento de extraer plantas, aproximadamente en un 13 % de los casos, saldrá la primera atacada en la tercera extracción.*

4.8.1 Características Numéricas

1. $E(X_g) = \dfrac{1}{p}$
2. $Var(X_g) = \dfrac{q}{p^2}$
3. $Des(X_g) = \sqrt{\dfrac{q}{p^2}}$

Para determinar cada una de ellas basta reemplazar en la definición de las características numéricas de una variable aleatoria discreta la $P(X_g=k)$.

4.9 Modelo Probabilístico Binomial Negativo

El Modelo Probabilístico Binomial Negativo describe experimentos que son una secuencia de Ensayos de Bernoulli independientes con parámetro **p**, hasta obtener **r** éxitos. Para describir este modelo se presenta el siguiente

Ejemplo 10: Sabiendo que la probabilidad de que una semilla esté afectada es 0.40 se desea calcular la probabilidad de realizar cinco extracciones para obtener tres semillas afectadas.

Solución:

El experimento ***E***: "Extraer semillas en forma sucesiva hasta obtener tres afectadas".
La variable X= Número de extracciones necesarias hasta que salgan 3 semillas afectadas.
$R(X)=3, 4, 5, ...$

Cuando un experimento consiste en repeticiones independientes de un Ensayo de Bernoulli hasta que ocurran **r** éxitos, se dice que se ajusta a un *Modelo Binomial Negativo*.

Definición 23: Los *parámetros* del Modelo Binomial Negativo son **r** y **p**, que representan el número de éxitos y la probabilidad de éxito respectivamente.

Definición 24: La *variable aleatoria Binomial Negativa* X_{bn} cuenta el *número de ensayos requeridos hasta* que *sale el suceso éxito r veces*.

El conjunto de valores que asume es $R(X_{bn})=r, r+1,....$

Definición 25: Sea X_{bn} una variable aleatoria con distribución binomial negativa. La probabilidad de que se realicen **k** ensayos hasta obtener **r** éxitos es

$$P(X_{bn} = k) = \binom{k-1}{r-1} p^r \, q^{k-r} \qquad \text{con } k=r, r+1, ... \qquad (4.5)$$

Retomando el Ejemplo 10 la probabilidad realizar cinco extracciones para obtener tres semillas afectadas.

$$P(X_{bn} = 5) = \binom{4}{2}(0.40)^3(0.60)^{5-3} = 0.13824$$

Este valor indica que *si se realiza muchas veces el experimento de extraer semillas, aproximadamente en un 14 % de los casos, saldrá la tercera semilla afectada en la quinta extracción.*

4.9.1. Características numéricas

1. $E(X_{bn}) = \dfrac{r}{p}$

2. $Var(X_{bn}) = \frac{r \bullet q}{p^2}$

3. $Des(X_{bn}) = \sqrt{\frac{r \bullet q}{p^2}}$

Para determinar cada una de ellas basta reemplazar en la definición de las características numéricas de una variable aleatoria discreta la $P(X_{bn}=k)$.

A continuación se muestra un cuadro resumen donde se presentan los modelos probabilísticos y sus características.

Tabla 3: Modelos Probabilísticos para variables Discretas y sus características

Modelo	*Parámetros del Modelo*	*Variable*	*Características Numéricas* *Esperanza*	*Varianza*
Bernoulli	p	X_{BER}: Número de éxitos.	$E(X_{BER})=p$	$Var(X_{BER})=p \bullet q$
Binomial	n y p	X_b: Número de éxitos en las n repeticiones.	$E(X_b)=n \bullet p$	$Var(X_b)= n \bullet p \bullet q$
Hipergeométrico	N, n y r	X_h: Número de éxitos en las n repeticiones.	$E(X_h)=n \bullet p$	$Var(X_b)=\frac{N-n}{N-1} \bullet n \bullet p \bullet q$ $p=r/N$
Poisson	λ	X_p: Número de éxitos en un cierto intervalo.	$E(X_p)=\lambda$	$Var(X_p)=\lambda$
Geométrico	p	X_g: Número de ensayos hasta que sale el primer éxito.	$E(X_g) = \frac{1}{p}$	$Var(X_g) = \frac{q}{p^2}$
Binomial Negativo	r y p	X_{bn}: Número de ensayos hasta que sale el r-ésimo éxito.	$E(X_{bn}) = \frac{r}{p}$	$Var(X_{bn}) = \frac{r \bullet q}{p^2}$

Ejercicios de Aplicación

1.

Se lanzan dos dados balanceados y homogéneos y se define la variable aleatoria "suma de los números aparecidos en las caras superiores". Hallar:

a) la distribución de probabilidades de la variable aleatoria definida.

b) el valor de: i) $P(X \leq 1)$. Ii) $P(X>3)$.

2.

Una urna contiene 5 fichas: 3 blancas y 2 azules. Se realiza el experimento “extraer una muestra de tamaño 3 con reposición” y se define la variable aleatoria “número de fichas blancas extraídas”.

a) Describir el espacio muestral asociado al experimento aleatorio.

b) Asociar a cada resultado posible del experimento el valor asignado por la variable aleatoria definida.

c) Hallar la distribución de probabilidades de la variable aleatoria.

d) Rehacer el inciso c) para el experimento “extraer una muestra de tamaño 3 sin reposición”, con la misma variable aleatoria.

3.

Para una inspección sanitaria se seleccionan 10 cerdos al azar de los criaderos de la zona mencionada en el Ejercicio 11 del Capítulo anterior. ¿Cuál es la probabilidad de que la mitad de los cerdos seleccionados tengan 16% o más de área pulmonar afectada por neumonía?.

a) ¿Cuál es el experimento aleatorio que describe el enunciado ? Indicar un elemento cualquiera del espacio muestral ***S***. Para dicho experimento:

i) ¿Cuál es el ensayo de Bernoulli y cuáles sus resultados posibles ?

ii) ¿Qué significa para este ejemplo que se repita un número fijo de veces el ensayo de Bernoulli ?

iii) ¿Qué significa la independencia de los ensayos?.

iv) ¿Cuál es la probabilidad del éxito? ¿Qué significa en esta situación que esta probabilidad se mantenga constante en la repetición de los ensayos?.

b) ¿Qué variable asociaría al experimento?

c) Calcular la probabilidad pedida, suponiendo que el experimento es binomial.

4.

A partir de la información suministrada en Ejercicio 8 del capítulo anterior se quiere hallar la probabilidad de que de un total de 6 descendientes de un cruzamiento de gatos heterocigotas se obtengan 2 con pelaje manchado.

a) ¿Cuál es el experimento aleatorio que describe el enunciado?. Detalle al menos dos elementos de ***S***. ¿Se ajusta este experimento a un Modelo Binomial?. ¿Por qué?.

b) ¿Cuál es la variable que debería asociarse al espacio muestral de tal forma que resulte una variable aleatoria binomial ?.

c) ¿Cuáles son los parámetros del modelo en esta situación ?. ¿Qué significan ?.

d) Hallar las características numéricas de la variable y dar su significado.

e) Calcular la probabilidad requerida.

5.

Se estima que el 90% de la cosecha de papas es buena encontrándose en estado de putrefacción las restantes, aunque esto no puede descubrirse a menos que las papas

se corten por la mitad. Si el estado de putrefacción no se contagia, ¿cuál es la probabilidad de que en una bolsa de 12 haya 9 en buen estado?. Justificar la respuesta

6.

La probabilidad de que un animal reciba una inyección de penicilina y sufra reacción desfavorable es de 0.10. Suponiendo que 9 animales reciben aplicaciones de este medicamento.

a) ¿Cuál es el experimento que se ajusta al modelo binomial y la variable aleatoria asociada?

b) Calcular la probabilidad de que:
i) todos los animales sufran reacción desfavorable.
ii) 6 sufran reacción desfavorable y 3 no.
iii) por lo menos 4 sufran reacción desfavorable.
iv) a lo sumo 2 sufran reacción favorable.

c) Realizar una representación gráfica de la distribución de probabilidades de la variable aleatoria.

d) Calcular las características numéricas de la variable, explicando qué información suministran para este problema. Establecer relaciones entre los valores de las características numéricas de la variable y el gráfico construido en el inciso c).

e) ¿Cuál es el número esperado de animales con reacción favorable?.

7.

Se afirma que una vacuna es eficaz en un 70%.

a) Hallar la probabilidad de que de 10 individuos que hayan recibido la vacuna:
i) 2 tengan la enfermedad.
ii) por lo menos 4 no tengan la enfermedad.
iii) 3 tengan la enfermedad y 7 no.
iv) a lo sumo 2 tengan la enfermedad.

b) Calcular $E(X_b)$ y $Var(X_b)$, dando su significado para este problema.

c) ¿Cuál es el número esperado de individuos para los cuales la vacuna no fue eficaz?.

8.

La probabilidad de que un niño esté afectado de Hymenolepis (enfermedad parasitaria) en ciertos barrios de la ciudad de Río Cuarto es de 0.5. A diez niños del sector mencionado se le realizan los análisis correspondientes y se determina “el número de niños infectados por Hymenolepis”.

a) ¿Se ajusta la experiencia al Modelo Binomial? ¿Cuál sería el supuesto del modelo binomial que podría no satisfacerse?. ¿Por qué?

b) Realizar una representación gráfica de la distribución de probabilidades de la variable aleatoria binomial correspondiente a esta situación.

c) Calcular la esperanza y la varianza de la variable, dando su significado para esta situación. Establecer relaciones entre las características numéricas y el gráfico de la distribución de probabilidades de la variable.

d) Suponiendo que, en lugar de 0.5, la probabilidad de que un niño esté afectado por el parásito es de 0.1, realizar un análisis similar al efectuado en los incisos anteriores.

e) Discutir las similitudes y/o diferencias entre los gráficos de las distribuciones de probabilidad construidos en uno y otro caso.

5 Variables Aleatorias Continuas

Objetivos:

- ♦ Identificar variables aleatorias continuas.
- ♦ Reconocer los parámetros de las distintas distribuciones.
- ♦ Establecer relaciones entre distintas distribuciones.

5.1 Variables Aleatorias Continuas

En el capítulo anterior se trabajó con variables aleatorias discretas. A continuación se estudiarán otro tipo de variables denominadas *variables aleatorias continuas*. Algunos ejemplos de este tipo de variables son:

- * tiempo de coagulación;
- * altura de las personas;
- * ganancia de peso de animales;
- * rendimiento de un cultivo;
- * tiempo de recuperación de cierta enfermedad;
- * errores de medición en experimentos científicos.

Estas variables, tienen la particularidad de asumir cualquier valor dentro de un cierto intervalo por lo que, a diferencia de las variables discretas, no es posible asociarles un valor de probabilidad puntual distinto de cero a cada valor de variable. Para aclarar ideas se muestra el siguiente

Ejemplo 1: Para estudiar el comportamiento del *peso* de animales recién nacidos de la raza Charolais se realizó el siguiente experimento aleatorio ***E***: “Se extrae al azar un animal recién nacido”.

Se sabe que todo experimento aleatorio tiene asociado un espacio muestral. En este caso no se pueden enumerar todos los elementos de dicho espacio (que se corresponden biunívocamente con los valores posibles de la variable peso) y por lo tanto, para construir la distribución de probabilidades de este tipo de variables se va a desarrollar otro método.

Generalmente, para representar las frecuencias de una variable de tipo continuo se utiliza una tabla de frecuencias agrupadas, lo que permite construir un Histograma. Éste consiste de una serie de rectángulos cuya base está dada por la longitud del intervalo y cuya altura es la frecuencia absoluta (como se presentó en el Capítulo 1).

Si se generan muestras de tamaño **n** de una variable aleatoria X, para cada una de ellas es posible construir la tabla de frecuencias agrupadas con intervalos $[\, l_{i-1}\,,\, l_i\,)$, de longitud $l_i - l_{i-1} = 1/n$. Es decir:

Intervalo	Frecuencia absoluta (f_i)
$I_1=[l_0, l_1)$	f_1
.	
.	
.	
$I_i=[l_{i-1}, l_i)$	f_i
.	
.	
.	
$I_n=[l_{n-1}, l_n]$	f_n

Si en el histograma asociado a la tabla se traza una poligonal que pase por los puntos (**c_i**, **f_i**), con **c_i** la marca de clase del intervalo I_i, a medida que aumenta el tamaño de la muestra, **n**, esa poligonal se hace más suave y se acerca cada vez más a un función **f**, tal como se muestra en el siguiente gráfico.

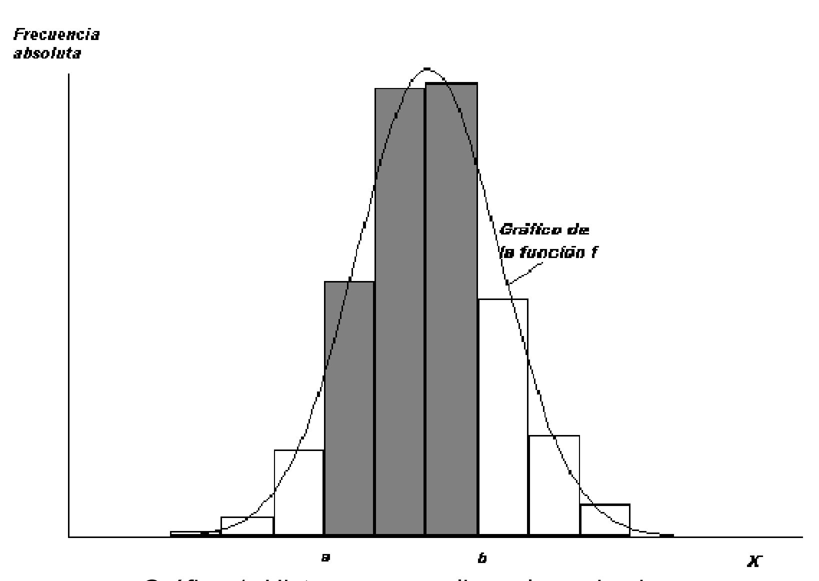


Gráfico 1: Histograma y poligonal suavizada.

Suponga establecida dicha función **f** y considere un intervalo [a,b] fijo. Se presentará a continuación la justificación intuitiva del cálculo de $P(a\leq X \leq b)$.

Si se consideran los intervalos **I_i** cuya unión da el [a,b], es posible ver que el área bajo la gráfica de **f** y por encima del intervalo [a,b] puede ser aproximada por la suma de las áreas de los rectángulos, **R_i** , de base el intervalo **I_i** y altura **f_i** . Es decir:

Área bajo **f** y por encima del intervalo [a,b] ≅
suma de las áreas de los rectángulos **R_i** (5.1)

Pero por otro lado, como se dijo, cada rectángulo tiene base de longitud **1/n** y altura **f_i** , de donde, su área es (base) • (altura)= (**1/n**) • (**f_i**) = frecuencia relativa del intervalo **I_i**.

Entonces (5.1) se puede reescribir como:

Area bajo **f** y por encima del intervalo [a,b] ≅
suma de las frecuencias relativas de los intervalos **I_i** (5.2)

pero esta última suma no es otra cosa que la frecuencia relativa correspondiente al intervalo [a,b], es decir:

Área bajo **f** y por encima del intervalo [a,b] $\cong$
frecuencia relativa del intervalo [a,b] (5.3)

Como se estableció en el capítulo 3, cuando el tamaño de la muestra crece, la frecuencia relativa se aproxima a la probabilidad, entonces :

Frecuencia relativa del intervalo [a,b] $\cong P(a \leq X \leq b)$ (5.4)

De (5.3) y (5.4), cuando **n** tiende a infinito, se deduce que:

Área bajo f y por encima del intervalo [a,b] = $P(a \leq X \leq b)$. (5.5)

La función **f** presentada se denomina función de densidad.

Definición 1: Una *función de densidad* es una función a valores reales ($f: \Re \rightarrow \Re$) si verifica las siguientes condiciones:

1. $f(x) \geq 0$, o sea que el gráfico de la función está por encima del eje de las abscisas **x**, para todos los valores de la variable.
2. El área bajo la curva función $f(x)$ y por encima del eje **x** es igual a 1.

En el Gráfico 2 se presentan ejemplos de funciones de densidad, si el área bajo la curva es 1.

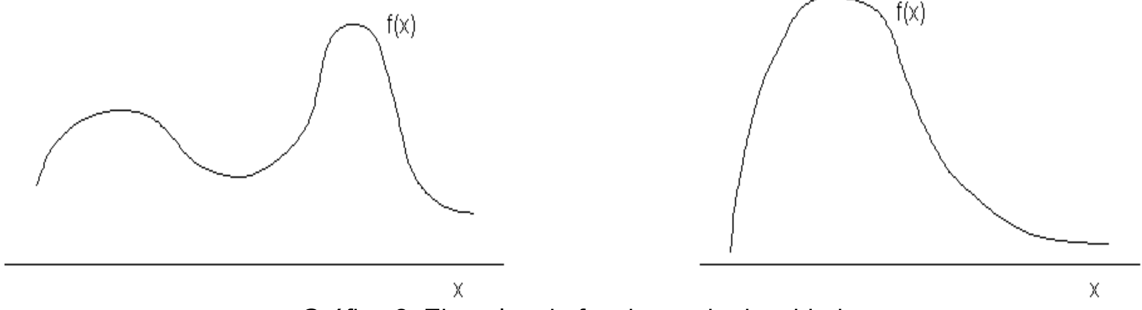


Gráfico 2: Ejemplos de funciones de densidad

Definición 2: Sea X es una variable aleatoria, es decir una función $X: \boldsymbol{S} \rightarrow \Re$. Se dice que X es una *variable aleatoria continua* si tiene asociada una función de densidad $f: \Re \rightarrow \Re$.

Por (5.5), la probabilidad de que la variable X tome valores entre **a** y **b**, $P(a \leq X \leq b)$, se calcula como el área limitada por la curva de la función de densidad $f(x)$, las rectas X=a, X=b y el eje **x**, tal como se puede observar en el Gráfico 3.

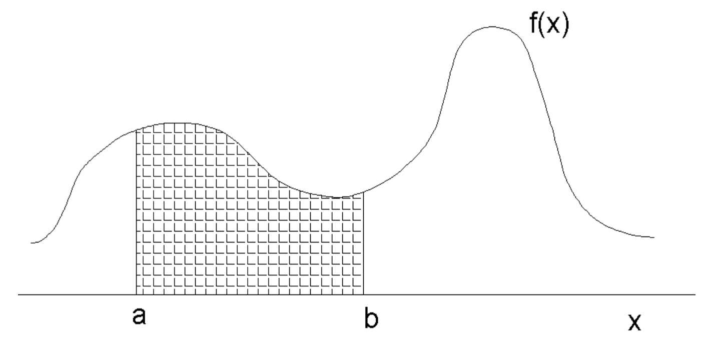


Gráfico 3: Área correspondiente a $P(a \leq X \leq b)$.

En base a esta definición, la probabilidad de que la variable aleatoria continua X tome un valor exactamente igual a **c** es cero, o sea $P(X=c)=0$. Esta probabilidad es igual al área encerrada por dos líneas verticales iguales, y como esa área no existe se puede decir que la probabilidad correspondiente a un *valor puntual es cero*.

Una forma de determinar el área es utilizando algunas herramientas del análisis matemático que no son muy simples, por lo cual las distribuciones de probabilidades de las variables aleatorias continuas más utilizadas fueron tabuladas. Por esto cada vez que se desea calcular una probabilidad se puede recurrir a la tabla correspondiente.

5.2 Características Numéricas de una variable aleatoria continua

Las características numéricas de las variables aleatorias continuas son, como en el caso discreto, la Esperanza, la Varianza y la Desviación Estándar.

Para determinar cada una de ellas son necesarias algunas herramientas matemáticas que no están al alcance de este texto, por lo tanto no se presentará la expresión general (como se hizo en el caso discreto). Para cada una de las distribuciones continuas presentadas a continuación se indicarán los valores de las características numéricas.

Así como hay varios modelos probabilísticos asociados con variables aleatorias discretas (Binomial, Poisson, Geométrico, etc.), hay varios modelos asociados con variables aleatorias continuas. En este texto sólo se presenta exhaustivamente el siguiente modelo.

5.3 Distribución Normal

La importancia de esta distribución radica en la enorme frecuencia con que aparece en todo tipo de situaciones de la vida cotidiana y también en el hecho de que juega un papel muy importante en la inferencia estadística clásica.

Karl F. Gauss, estudiando la distribución de los errores, resultantes de medir reiteradamente una misma magnitud, probó que seguían esta distribución, por lo cual es conocida como distribución Normal, de Gauss o gaussiana. La apariencia gráfica de la distribución normal es una curva simétrica en forma de campana (campana de Gauss).

Un gran número de estudios indica que la distribución normal proporciona una adecuada representación, por lo menos en una primera aproximación, de las distribuciones de una gran cantidad de variables físicas. Algunos ejemplos específicos incluyen datos meteorológicos tales como temperatura, mediciones efectuadas en organismos vivos, mediciones físicas realizadas en partes manufacturadas, errores de instrumentación, etc. . A continuación se da la definición formal de esta distribución.

Definición 3: Una variable aleatoria continua X se dice que tiene *distribución de probabilidades normal si* su función de densidad se describe por:

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right) \qquad (5.6)$$

donde $\mathbf{x}$ y μ son números reales cualesquiera y σ es un número real positivo.

La representación gráfica de esta función de densidad se muestra a continuación:

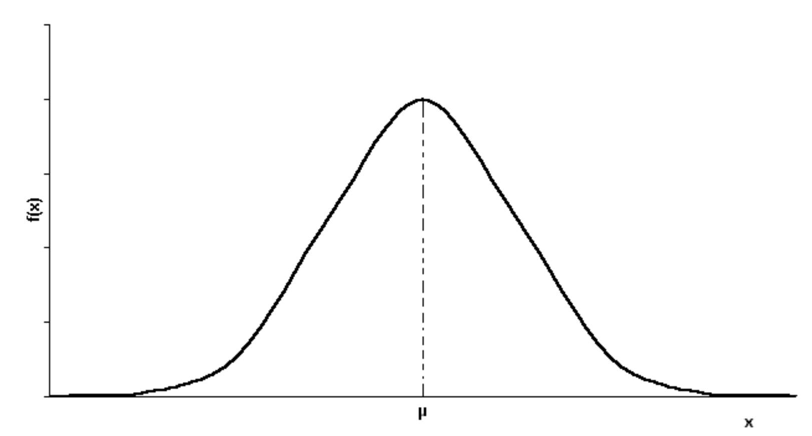


Gráfico 4: Función de densidad correspondiente a una variable con distribución normal

Definición 4: Los *parámetros* de la distribución Normal son μ y σ.

Como μ y σ pueden tomar infinitos valores existen infinitas distribuciones normales. Estos parámetros influyen en el gráfico de la función de densidad de la siguiente manera:

- μ es el punto sobre el eje x por donde pasa el eje de simetría de la curva, luego se verifica que $P(-\infty \leq X \leq \mu) = P(\mu \leq X \leq \infty) = 0.50$; es decir la probabilidad de que la variable X tome valores inferiores o iguales a μ es igual a la probabilidad de que X tome valores superiores o iguales a μ y ambas son iguales a 0.50.
- σ determina la forma de la curva, en cuanto a la agudeza de la misma.

Una relación importante entre los parámetros es:

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6826$$
$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9546$$
$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9973$$

Esta relación también puede expresarse de la siguiente manera:

$\mu \pm \sigma$ contiene el 68.26% de los valores de la variable
$\mu \pm 2\sigma$ contiene el 95.46% de los valores de la variable
$\mu \pm 3\sigma$ contiene el 99.73% de los valores de la variable

En cada uno de los siguientes gráficos se muestran distribuciones normales con distintos valores para los parámetros.

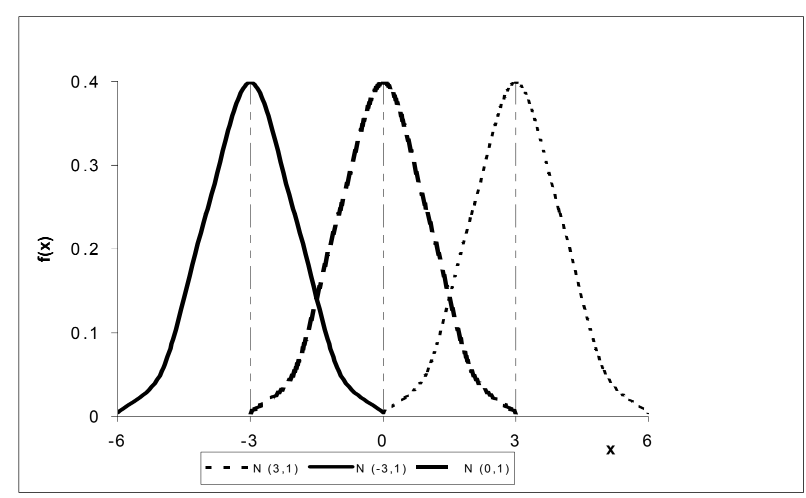


Gráfico 5 (a): Funciones de densidad correspondientes a variables aleatorias con distribución normal, con diferentes μ e igual σ.

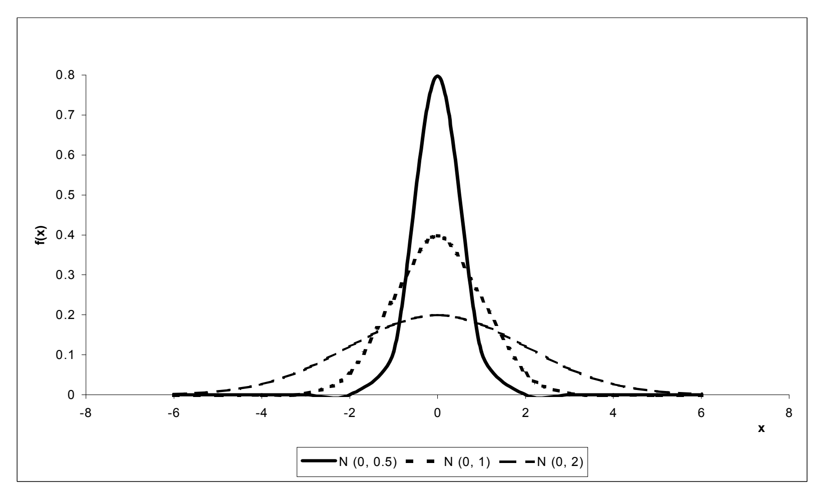


Gráfico 5 (b): Funciones de densidad correspondientes a variables aleatorias con distribución normal, con igual μ y distinto σ.

Para indicar que la variable X tiene distribución Normal, con parámetros μ y σ, se escribe $X \sim N(\mu ,\sigma)$.

Las *características numéricas* de una variable aleatoria con distribución normal de parámetros μ y σ son

$$E(X)=\mu \ , \ Var(X)=\sigma^2 \ y \ Des(X)=\sigma$$

Éstas coinciden con los parámetros de dicha distribución.

Ejemplo 2: Retomando el Problema 1.1, se está interesado en estudiar el peso de novillos de una cierta edad de la raza Charolais a fin de analizar la posibilidad de suministrarles una nueva dieta rica en proteínas. Se sabe que la variable peso tiene distribución normal de parámetros μ=300 kg. y σ=20 kg., es decir $X \sim N(300,20)$.

Dado que los parámetros de la distribución normal coinciden con las características numéricas de la variable, se tiene que $E(X)=300$ y $Des(X)=20$. El significado de cada uno de estos valores es:

- $E(X)=300$ kg. , indica que los animales pesan alrededor de 300 kg. y que la cantidad de animales con peso muy alejado de 300 kg. (mayores o menores) es baja.

- $Des(X)=20$ kg. significa que los pesos de los animales se alejan, en promedio, 20 kg. del peso medio (300 kg.).

La probabilidad de que un animal pese exactamente 320 kg. es 0, lo cual no significa que sea imposible que eso ocurra, sino que entre los infinitos valores de pesos que existen es muy difícil encontrar un animal que pese exactamente 320 kg.

El parámetro $\mu=300$ de la distribución normal indica que la probabilidad de que un animal pese más de 300 kg. es igual a la probabilidad de que pese menos de 300 kg. y vale 0.50. En símbolos $P(X \geq 300) = P(X \leq 300) = 0.50$.

5.3.1 Cálculo de probabilidades de una variable con distribución normal

Para ver como calcular probabilidades de una variable aleatoria continua con distribución normal se retomará el Ejemplo 2.

Ejemplo 3: Se está interesado en determinar la probabilidad de que un novillo de la raza Charolais elegido al azar tenga un peso inferior a 320 kg., es decir $P(X < 320)$.

Antes de entrar en detalles acerca de como obtener la probabilidad asociada a una variable aleatoria normal se va a definir una distribución normal muy especial y de gran utilidad para resolver el problema de calculo de probabilidades.

5.3.1.1 Distribución Normal Estándar

Dado que para cada par de valores de parámetros se genera una distribución normal, parecería necesario contar con infinitas tablas de dicha distribución. Sin embargo, con una sola tabla se pueden calcular probabilidades, pues haciendo una transformación cualquier variable aleatoria con distribución normal puede convertirse en otra variable aleatoria con distribución normal de parámetros $\mu=0$ y $\sigma=1$. Esta distribución recibe el nombre de *Distribución Normal Estándar* y es la que está tabulada. En el siguiente gráfico se muestra su función de densidad.

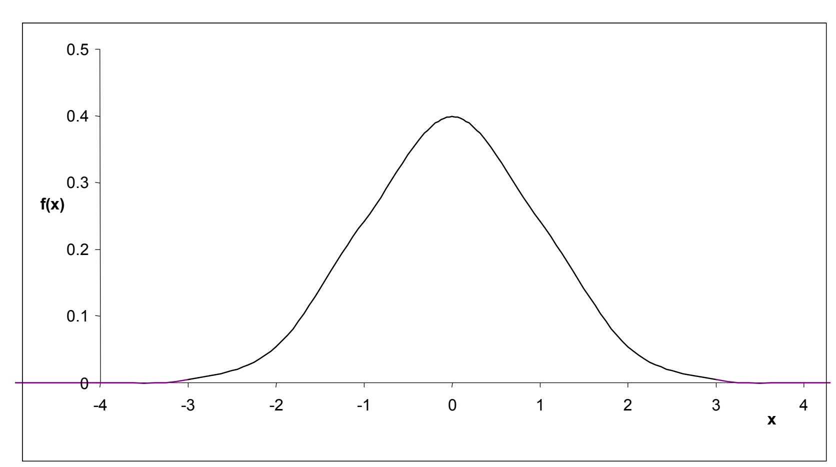


Gráfico 6: Función de densidad de la distribución Normal Estándar

5.3.1.2 Estandarización

Para convertir una variable aleatoria $X \sim N(\mu,\sigma)$ en una variable aleatoria normal estándar se realiza la siguiente transformación

$$Z = \frac{(X-\mu)}{\sigma}$$

A partir de ella surgen resultados importantes

1. La variable Z tiene distribución normal.
2. Los parámetros de esa distribución normal son $\mu=0$ y $\sigma=1$.

El primer resultado es una consecuencia del siguiente resultado general: "Una función lineal de una variable aleatoria con distribución normal es una variable aleatoria con distribución normal" (Meyer, P. 1992).

El segundo resultado se puede probar fácilmente teniendo en cuenta que los parámetros coinciden con las características numéricas y usando las propiedades de Esperanza y Varianza, como se indica a continuación.

Sea $$Z = \frac{X-\mu}{\sigma} \qquad (5.7)$$

a) tomando esperanza a ambos miembros en la expresión (5.7) se tiene

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma} \bullet E(X-\mu)$$ (por E_2)

$$= \frac{1}{\sigma} \bullet [E(X) - E(\mu)]$$ (por E_3)

$$= \frac{1}{\sigma} \bullet (\mu - \mu)$$ (por E_1 y dado que $E(X)=\mu$)

$$= \frac{1}{\sigma} \bullet 0 = 0$$

b) tomando varianza a ambos miembros en la expresión (5.7) se tiene

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} \bullet Var(X-\mu)$$ (por V_2)

$$= \frac{1}{\sigma^2} \bullet [Var(X) + Var(\mu)]$$ (por V_3)

$$= \frac{1}{\sigma^2} \bullet (\sigma^2 - 0)$$ (por V_1 y dado que $Var(X)=\sigma^2$)

$$= \frac{1}{\sigma^2} \bullet \sigma^2 = 1$$

entonces $Var(Z)=1$ y por lo tanto $Des(Z)=1$.

5.3.1.3 Manejo de la tabla de la distribución Normal Estándar

La Tabla C del Apéndice muestra la probabilidad acumulada hasta un cierto valor positivo de abscisa z, $P(Z\leq z)$.

Solución para el Ejemplo 3:
Recordando que se desea determinar $P(X<320)$, que corresponde gráficamente al área sombreada en el Gráfico 7

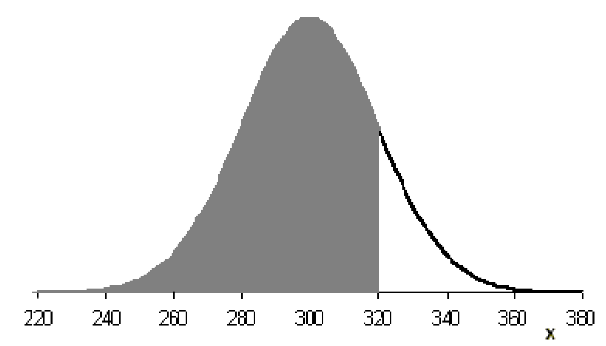


Gráfico 7: Probabilidad de que el peso de un animal sea inferior a 320 kg.

Lo primero a tener en cuenta es que la variable en estudio es $X \sim N(300,20)$, luego para determinar la probabilidad deseada se debe estandarizar, o sea definir la variable

$$Z = \frac{X-300}{20} \sim N(0,1)$$

Luego $P(X<320) = P\left(\frac{X-300}{20} < \frac{320-300}{20}\right) = P(Z<1) = 0.8413$.

Al estandarizar el gráfico anterior se transformó en el siguiente

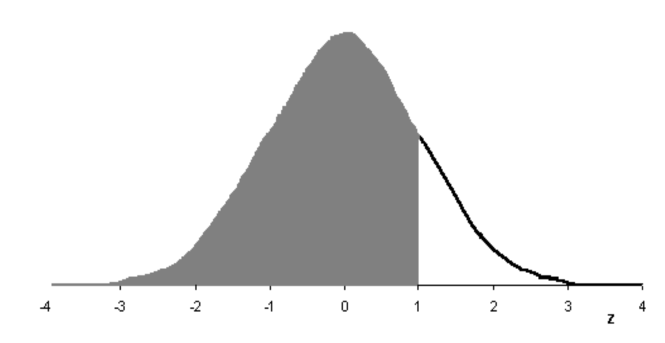


Gráfico 8: Probabilidad de que la variable estandarizada sea inferior a 1.

Cabe destacar que las áreas representadas en ambos gráficos son equivalentes.

El valor de la probabilidad indica que el 84.13% de los animales pesan menos de 320 kg.

Ejemplo 4: Bajo los supuestos del Ejemplo 2, se desea determinar la probabilidad de que el peso de un novillo de la raza Charolais
a) sea superior a 340 kg. b) esté entre 285 kg. y 350 kg. c) sea inferior a 220 kg.

Solución:
Como la variable en estudio $X \sim N(300,20)$, las probabilidades son:

a)
$$P(X > 340) = P\left(\frac{X-300}{20} > \frac{340-300}{20}\right) = P(Z > 2) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$$
b)
$$P(285 < X < 350) = P\left(\frac{285-300}{20} < \frac{X-300}{20} < \frac{350-300}{20}\right) = P(-2.75 < Z < 2.5) =$$
$$= P(Z < 2.5) - P(Z < -2.75) = 0.9938 - 0.0030 = 0.9908$$
donde $P(Z < -2.75)=P(Z > 2.75)=1-P(Z < 2.75)$

c)
$$P(X < 220) = P\left(\frac{X-300}{20} < \frac{220-300}{20}\right) = P(Z < -4) = 0$$

Gráficamente la probabilidad $P(X>340)$ es

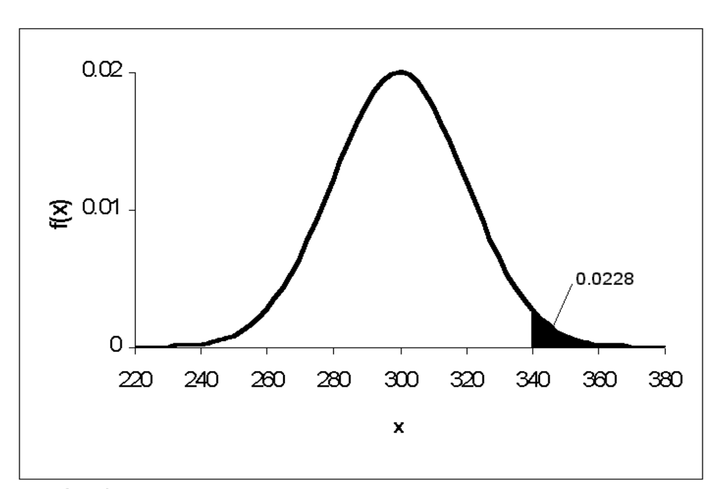


Gráfico 9: Probabilidad de que el peso de un animal sea superior a 340 kg.

Muchas variables aleatorias de tipo continuo se ajustan a otras distribuciones tales como la Distribución Uniforme, Gamma, Beta, Weibull y Exponencial, pero ellas no serán motivo de discusión de este texto. Sí se definirán otras distribuciones de tipo continuo muy útiles para la estadística inferencial.

5.4 Otras distribuciones continuas

Las distribuciones continuas que serán definidas a continuación son muy utilizadas en la teoría estadística. Además tienen la particularidad de que sus parámetros son llamados *grados de libertad*, los cuales indican intuitivamente la cantidad de información independiente con la que se cuenta.

Por ejemplo, si se tiene una muestra de tamaño $\mathbf{n}$, la cual puede ser denotada por $(X_1,X_2,...,X_n)$, y $\overline{X}$ es su media. Si ahora se tiene la suma de los desvíos $\sum_{i=1}^{n}\left(X_i - \overline{X}\right)$, se ve que este valor depende sólo de $\mathbf{n-1}$ (de los $\mathbf{n}$) sumandos y el restante esta unívocamente determinado, por esto se dice hay $\mathbf{n-1}$ términos independientes y cuando se habla de los grados de libertad asociados a una distribución, se lo hace en estos términos.

5.4.1 Distribución Ji-Cuadrado (χ^2)

Definición 5: La variable aleatoria continua Y tiene *distribución de probabilidades Ji-Cuadrado con n grados de libertad* ($Y \sim \chi^2_n$), si puede expresarse como la suma de los cuadrados de n variables aleatorias independientes normales estándar, o sea

$$Y = \sum_{i=1}^{n} X_i^2 \qquad \text{donde } X_i \sim N(0,1) \text{ independientes, } i=1,2,...,n$$

Definición 6: El *parámetro* de la distribución χ^2_n es $\mathbf{n}$, o sea sus grados de libertad.

Notar que en la Definición 5 aparece un nuevo concepto *variables aleatorias independientes* y que la Definición 6 relaciona directamente los grados de libertad de la distribución Ji-Cuadrado con la cantidad de variables aleatorias independientes que intervienen en su definición. Falta decir entonces cuando $\mathbf{n}$ variables aleatorias serán consideradas independientes, por lo cual se da la siguiente :

Definición 7: Sea $\mathbf{E}$ un experimento aleatorio y S el espacio muestral asociado a él, se dice que las variables aleatorias $X_1, X_2, \ldots, X_n$ definidas sobre S son independientes, si para cualquier intervalo $[a_1,b_1],[a_2,b_2],....,[a_n,b_n]$, se verifica

$$P(a_1 < X_1 < b_1,...,a_n < X_n < b_n) = P(a_1 < X_1 < b_1)\cdots P(a_n < X_n < b_n)$$

La función de densidad de una variable aleatoria con distribución χ^2_n es una función $f:\Re^+ \to \Re^+$ cuya forma explícita es

$$f(x) = \frac{x^{n/2}\exp(x/2)}{2^n\,\Gamma(n)}$$

donde $\Gamma(n)$ es la función Gamma.(Meyer P. 1992).

El gráfico de la función de densidad de una variable aleatoria con distribución Ji-Cuadrado depende del parámetro o sea de los grados de libertad. A continuación se muestra el gráfico de la función de densidad para $n=1$, $n=2$, $n\geq 3$.

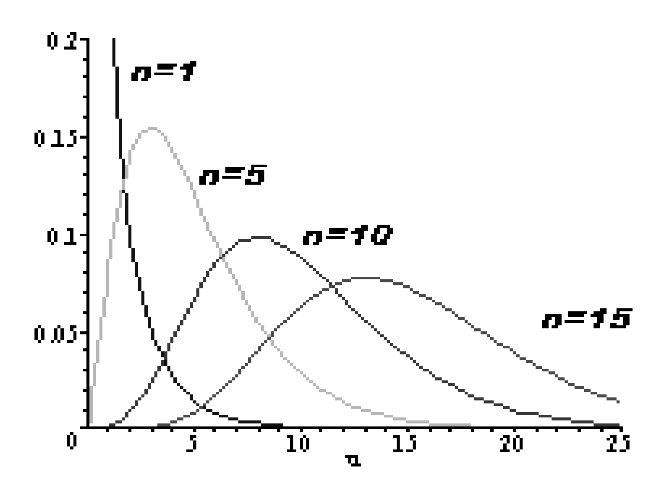


Gráfico 10: Funciones de densidad de la distribución
Ji-Cuadrado de acuerdo a sus grados de libertad

Las *características numéricas* de una variable aleatoria Ji-Cuadrado con $\mathbf{n}$ grados de libertad son:

$$E\left(\chi_n^2\right) = n \quad \text{y} \quad Var\left(\chi_n^2\right) = 2n$$

Esta distribución verifica la propiedad reproductiva, o sea, si X e Y son dos variables aleatorias independientes que tienen distribuciones χ_n^2 y χ_m^2 respectivamente, entonces la variable suma $X+Y$ tiene una distribución Ji-Cuadrado con $\mathbf{n+m}$ grados de libertad.(Meyer, P. 1992).

5.4.2 Distribución t de Student (t)

Definición 8 : La variable aleatoria continua T tiene *distribución de probabilidades t de Student con n grados de libertad* ($T \sim t_n$) si puede expresarse como el cociente de dos variables aleatorias independientes X (normal estándar) e Y (raíz cuadrada de una Ji-Cuadrado dividida sus grados de libertad), o sea

$$T = \frac{X}{Y}, \quad \text{donde } X \sim N(0,1) \quad \text{e} \quad Y = \sqrt{\frac{\chi_n^2}{n}}.$$

Definición 9: El *parámetro* de la distribución t_n es $\mathbf{n}$, o sea los grados de libertad.

Notar que el parámetro depende de los grados de libertad de la variable aleatoria Y.

La función de densidad de la variable aleatoria T con distribución t_n es una función $f:\Re \rightarrow \Re^+$ cuya forma explícita es

$$f(x) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2} \quad , -\infty < x < \infty$$

El gráfico de la función de densidad de una variable aleatoria con distribución t-Student

depende del parámetro o sea de los grados de libertad. A continuación se muestra el gráfico de la función de densidad para algunos valores de **n**.

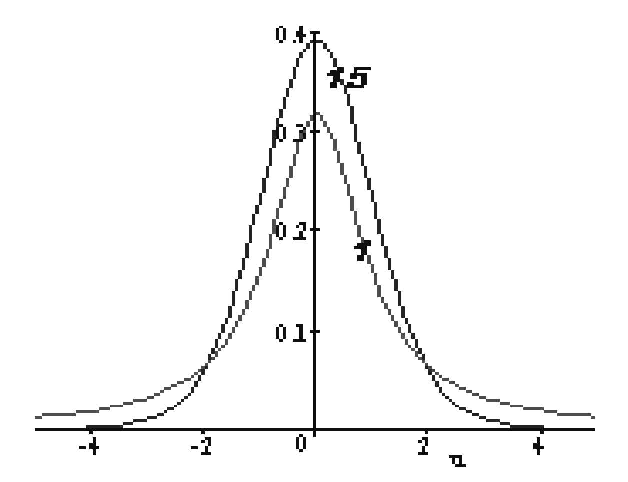


Gráfico 11: Función de densidad de la distribución t de Student
de acuerdo a sus grados de libertad

La función de densidad de una variable con distribución t de Student es simétrica respecto de cero y, comparada con la distribución normal estándar, tiene las colas más pesadas aproximándose a la función de densidad de ésta a medida que los grados de libertad aumentan.

Las *características numéricas* de una variable aleatoria T con distribución t de Student con **n** grados de libertad son

$E(t_n) = 0$, para $n > 1$ $\qquad$ $Var(t_n) = \dfrac{n}{n-2}$, para $n \geq 3$

5.4.3 Distribución F de Fisher (F)

Definición 10: Una variable aleatoria continua F *tiene distribución de probabilidades F de Fisher con n y m grados de libertad* ($F \sim F_{n,m}$) si puede expresarse como el cociente entre dos variables aleatorias independientes, X e Y (con distribuciones χ^2_n y χ^2_m respectivamente) dividido sus grados de libertad, o sea

$$F = \frac{X/n}{Y/m}, \qquad \text{donde } X = \chi^2_n \text{ e } Y = \chi^2_m$$

Definición 11: Los *parámetros* de la distribución F de Fisher son **n** y **m**, o sea los grados de libertad del numerador y del denominador, respectivamente.

La función de densidad de una variable aleatoria con distribución $F_{n,m}$ es tal que $f{:}\Re^+ \rightarrow \Re^+$ y cuya forma explícita es

$$f(x) = \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2} x^{n/2-1} \left(1 + \frac{n}{m}x\right)^{-(n+m)/2}, \qquad x > 0$$

El gráfico de la función de densidad se presenta a continuación para algunos grados de libertad.

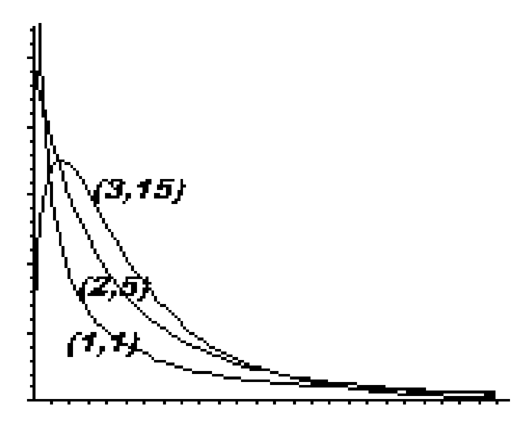


Gráfico 12: Funciones de densidad de la distribución F de Fisher de acuerdo a sus grados de libertad

Las *características numéricas* de una variable aleatoria **F** con distribución F de Fisher con **n** y **m** grados de libertad son:

$E(F) = \frac{m}{m-2}$, para $m > 2$

$Var(F) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$, para $m > 4$

Las distribuciones Ji-Cuadrado, t de Student y F de Fisher definidas en esta sección son comúnmente llamadas distribuciones centrales. La definición de las distribuciones no centrales correspondientes pueden ser consultadas en Johnson, N.L. y Kotz, S. (1970). Por otra parte estas distribuciones pueden aproximarse a la Normal a medida que aumentan los grados de libertad, lo que puede ser observado en los Gráficos 13, 14 y 15.

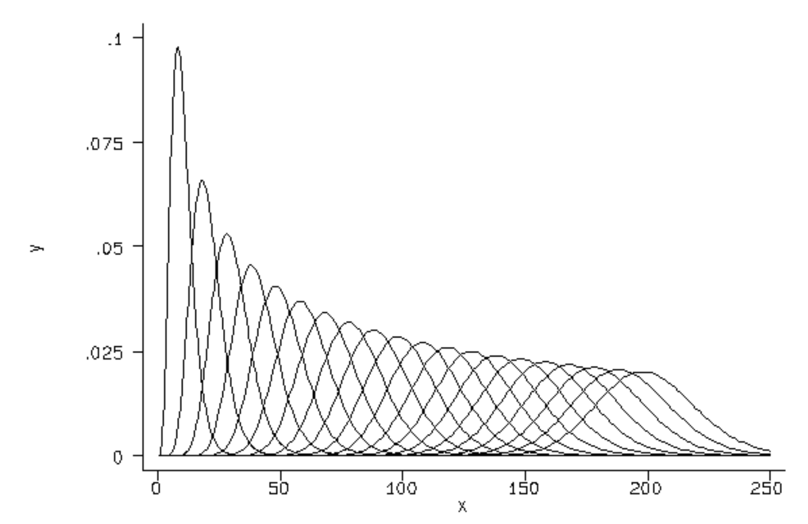


Gráfico 13: Funciones de densidad de la distribución Ji-Cuadrado con grados de libertad n=10,...,200

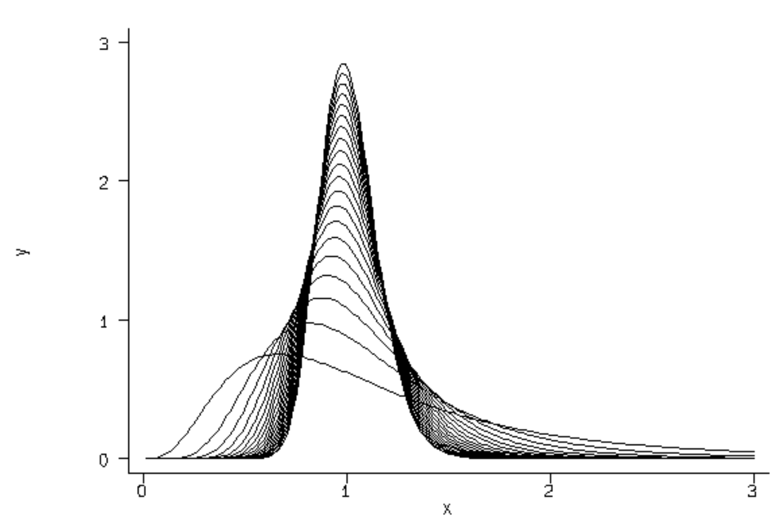


Gráfico 14: Funciones de densidad de la distribución F con grados de libertad (10,10), ..., (200,200)

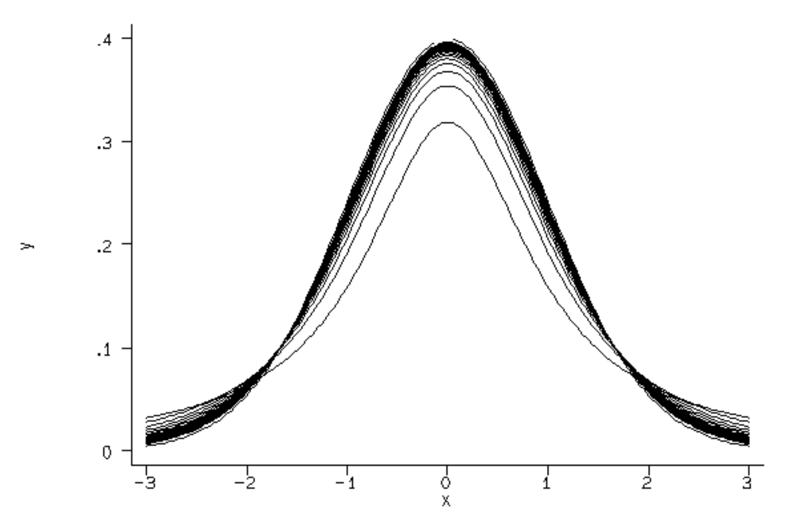


Gráfico 15: Funciones de densidad de la distribución t con grados de libertad 1, ...,30

5.4.4 Uso de Tablas

Como en los casos anteriores existen tablas que permiten calcular probabilidades.

- La Tabla E del Apéndice (distribución χ^2) presenta el valor de abscisa **z** que corresponde a un valor del parámetro (grados de libertad) y a la probabilidad de los valores de variable mayores a **z**, $P(\chi^2_n \geq z)$.
- La Tabla D del Apéndice (distribución **t**) presenta el valor de abscisa **z** que corresponde a un valor del parámetro (grados de libertad) y a la probabilidad acumulada hasta el valor de **z** ,.$P(t_n \leq z)$.
- La Tabla F del Apéndice (distribución **F**) presenta el valor de abscisa **z** que corresponde a valores de los parámetros (grados de libertad n_1 y n_2) y a la probabilidad $P(F_{n1,n2} \geq z)$.

5.5 Teoría elemental del muestreo

La teoría del muestreo estudia la relación entre una población y las muestras tomadas de ella. Por ejemplo para la *estimación* de magnitudes desconocidas de una población tales como media y varianza, llamadas a menudo parámetros de la población, se utilizan las características numéricas de la muestra, llamadas estadísticos. Entonces se puede decir que cuando se estiman valores poblacionales a través de una muestra se dice que se está haciendo Inferencia, en el capítulo siguiente se volverá a discutir sobre este tema.

Se puede decir, entonces, que la *base de la inferencia es la muestra*, por lo que el concepto dado a continuación es muy importante.

5.5.1 Muestras Aleatorias

Para entender este concepto se trabajará sobre el siguiente

Ejemplo 5: Para el Problema 1.1, en el cual el objetivo es estudiar el efecto de una nueva dieta sobre novillos de la raza Charolais a través de la variable peso, se tiene:

E: "extraer un novillo al azar de la raza Charolais".

Si se realizan **n** repeticiones de este experimento la muestra de unidades es

$$(animal_1, animal_2,..., animal_n)$$

Si a cada uno de estos **n** animales se les suministra la dieta y luego se registra su peso se obtiene, por ejemplo, la siguiente muestra estadística

$$(490, 530, 510, ..., 470)$$

donde

490 es el peso del $animal_1$ de la muestra de animales seleccionada,
530 es el peso del $animal_2$ de la muestra de animales seleccionada,
510 es el peso del $animal_3$ de la muestra de animales seleccionada,
.
.
.
470 es el peso del $animal_n$ de la muestra de animales seleccionada.

Si se toma otra muestra de **n** animales de la misma población se obtiene otra muestra estadística, por ejemplo

$$(480, 500, 540, ..., 450)$$

donde

480 es el peso del $animal_1$ de la muestra de animales seleccionada,
500 es el peso del $animal_2$ de la muestra de animales seleccionada,
540 es el peso del $animal_3$ de la muestra de animales seleccionada,
.
.
.
450 es el peso del $animal_n$ de la muestra de animales seleccionada.

Las dos muestras estadísticas indicadas anteriormente se obtuvieron *después* de realizar el experimento. Si se deseara indicar los valores de peso de **n** animales *antes* de realizar el experimento se los debería representar como

$$(X_1, X_2, ..., X_n)$$

donde:

X_1: representa el peso del animal que será elegido en primer lugar.
X_2: representa el peso del animal que será elegido en segundo lugar.
X_3: representa el peso del animal que será elegido en tercer lugar.
.
.
.
X_n: representa el peso del animal que será elegido en el lugar **n**.

$X_1, X_2, ..., X_n$ son **n** variables aleatorias independientes e idénticamente distribuidas, esto es, el peso de un animal no influye en el de otro y todas tienen la misma distribución que la variable X.

Definición 12: Sea X una variable aleatoria con cierta distribución en probabilidades. Sean, $X_1, X_2, ..., X_n$, **n** variables aleatorias independientes e idénticamente distribuidas. Entonces a $(X_1, X_2, ..., X_n)$ se la llama *muestra aleatoria* de tamaño **n** de la variable aleatoria X.

Los valores numéricos de la variable en estudio una vez realizado el experimento, se denotan $(x_1, x_2, ..., x_n)$ que es la *muestra estadística* correspondiente al experimento realizado.

Ahora bien, si $(X_1, X_2, ..., X_n)$ es una muestra aleatoria entonces los estadísticos resultan variables aleatorias ya que son funciones de **n** variables aleatorias y por tanto tendrán asociados una distribución de probabilidades.

A continuación se presentan algunos resultados básicos que permiten obtener las distribuciones de probabilidades de los estadísticos más utilizados para estimar valores poblacionales: la media y la varianza muestral

5.5.2 Distribución de probabilidades de la media muestral

Sea X una variable aleatoria con $E(X)=\mu$ y $Var(X)=\sigma^2$. Sea $\overline{X}$ el promedio muestral de una muestra aleatoria de tamaño **n**, entonces:

a) $E(\overline{X}) = \mu$

b) $Var(\overline{X}) = \dfrac{\sigma^2}{n}$

c) Para n grande ($n \geq 30$), $\overline{X} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$. Esta última expresión es un abuso de notación pues en realidad $\overline{X}$ *tiende* a la distribución Normal.

Los incisos a) y b) se pueden demostrar fácilmente, utilizando las propiedades de Esperanza y Varianza dadas en la Sección 4.3.1, en tanto que el punto c) es lo que se conoce como Teorema Central del Límite (Mendenhall, W. et. al. 1994). Este último resultado es de gran relevancia pues asegura que cualquiera sea la distribución de la variable X para una muestra de tamaño considerable, la *distribución de la media muestral,* $\overline{X}$*, es aproximadamente Normal.*

Cuando la variable X tiene distribución Normal, la distribución de la media muestral es Normal

5.5.3 Distribución de probabilidades de la varianza muestral

Así como $\overline{X}$ resulta una variable aleatoria con una cierta distribución, también S^2 es una variable aleatoria y su distribución surge del siguiente resultado:

Sea X una variable aleatoria con $E(X)=\mu$ y $Var(X)=\sigma^2$. Sean $\overline{X}$ y S^2 la media y la varianza muestral respectivamente, entonces

a) $E(S^2)=\sigma^2$

b) Si además X tiene distribución normal entonces $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.(Mendenhall, W. et. al. 1994)

Una aplicación de estos dos conceptos se presenta en el siguiente

Ejemplo 6: Bajo los supuestos del Ejemplo 2, se selecciona una muestra de $n=25$ novillos de esa raza con el objeto de determinar las siguientes probabilidades:

a) el peso promedio sea inferior a 295 kg.
b) la varianza del peso de los animales sea inferior a 324 kg.^{2}.

Solución:
Para este problema $X \sim N(300,20)$.

a) La variable de interés en este caso es la media muestral, la cual se distribuye de la siguiente manera $\overline{X} \sim N\left(300, 20/\sqrt{25}\right)$. Luego para determinar lo deseado se procede así

$$P\left(\overline{X} < 295\right) = P\left(\frac{\overline{X}-300}{20/\sqrt{25}} < \frac{295-300}{20/\sqrt{25}}\right) = P(Z < -1.25) = 0.1056 .$$

Si se seleccionara muchas veces grupos de 25 animales aproximadamente en el 10.56% de las veces el peso promedio será menor a 295 kg.

b) La variable de interés en este caso es la varianza muestral la cual, multiplicada por ciertas constantes se distribuye de la siguiente manera $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. Luego para calcular la probabilidad pedida se procede así.

$$P\left(S^2 < 324\right) = P\left(\frac{(n-1)}{\sigma^2} \bullet S^2 < \frac{(n-1)}{\sigma^2} \bullet 324\right) = P\left(\chi^2_{24} < \frac{24}{400} \bullet 324\right) = P\left(\chi^2_{24} < 19.44\right) \cong 0.25$$

Si se seleccionara muchas veces grupos de 25 animales aproximadamente en el 25% de las veces la varianza de los pesos será inferior a 324 kg.^{2}.

5.6 Relación entre Binomial, Poisson y Normal

Una propiedad importante de la distribución normal es que puede aproximar a la distribución Binomial y Poisson.

Se ha demostrado que si $X_b \sim B(n,p)$ con **n** grande ($n>30$) y **p** no muy cercano a 0 o a 1, entonces la variable

$$\frac{X_b - np}{\sqrt{n \bullet p \bullet q}}$$

se aproxima a la distribución normal estándar (Este resultado es válido debido al Teorema Central del Límite ya citado). Así se puede calcular $P(a \le X_b \le b)$ considerando a $X_b \sim N(n \bullet p, \sqrt{n \bullet p \bullet q})$.

Análogamente la distribución normal puede aproximar a una distribución de Poisson cuando $\lambda>5$, es decir $X_p \sim N(\lambda, \sqrt{\lambda})$.

Ejercicios de Aplicación

1.

Sea X una variable aleatoria continua con distribución normal, con media cero y varianza 1. En símbolos X ~ N(0,1). Hallar y graficar:

a) P(X < -1.96) ***b)*** P(X > 0) ***c)*** P(X > 2.45)

d) P(-1.64 < X < 1.64) ***e)*** P(X>-6.5)

2.

Rehacer el ejercicio anterior bajo el supuesto de que X ~ N(3,2).

3.

Bajo el supuesto de que X ~ N(0,1), determinar el valor de a tal que:

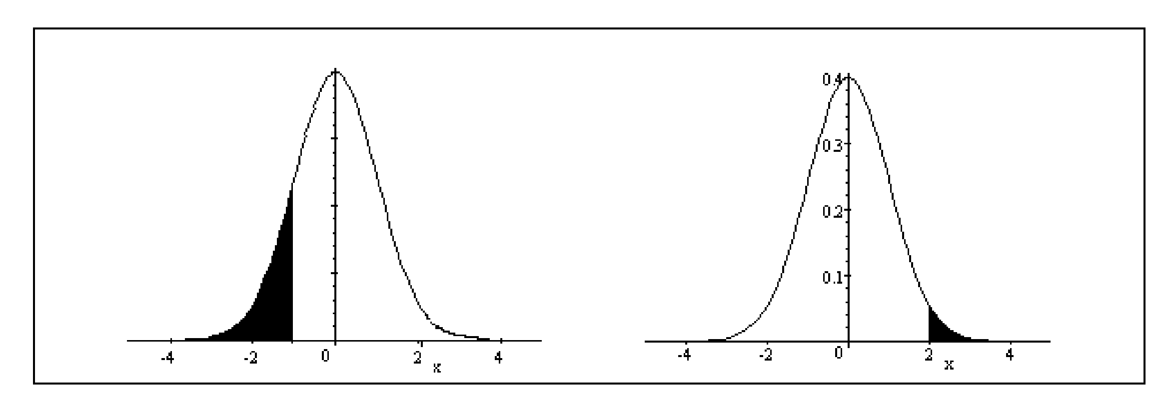
a) P(X < a) = 0.025 ***b)*** P(X > a) = 0.975 ***c)*** P(X < a) = 0.90

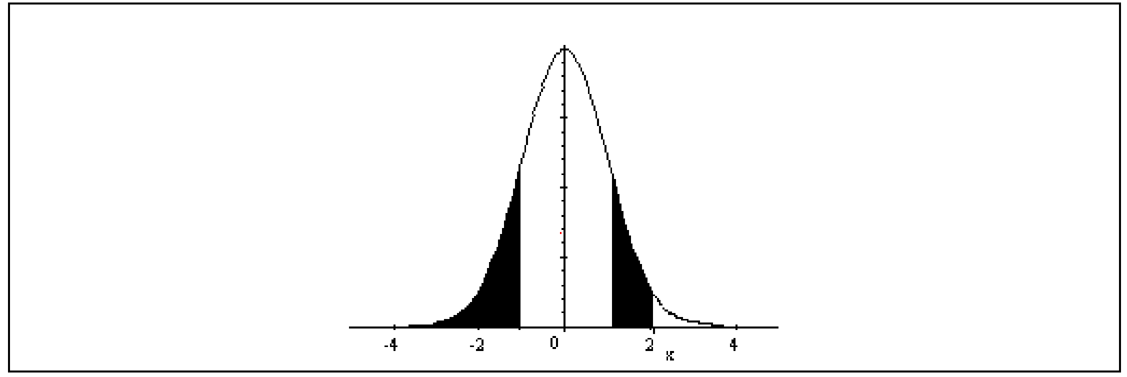
d) P(X > a) = 0.10 ***e)*** P(X < a) = 0.6844 ***f)*** P(X < a) = 0.1075

g) P(X > a) = 0.0668 ***h)*** P(X > a) = 0.8554 ***i)*** P(X > a) = 0.20

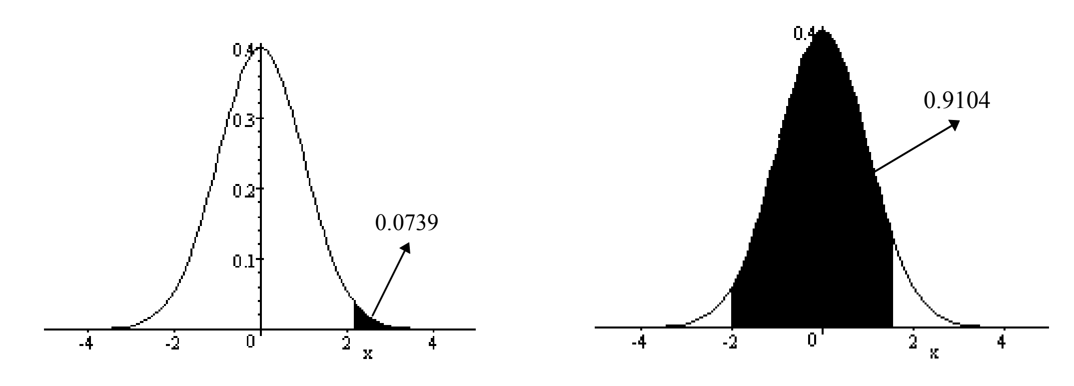
4.

a) Encontrar el valor del área sombreada debajo de las siguientes funciones de densidad correspondientes a una variable aleatoria normal estándar.



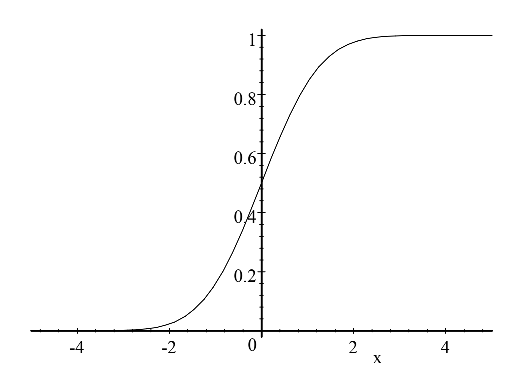


b) Hallar el valor de x que corresponde a la probabilidad indicada en los siguientes gráficos:



5.

Dada una variable aleatoria se define como función de distribución a la probabilidad de que la variable tome valores menores a un valor dado. En símbolos $F(x)=P(X<x)$. El siguiente gráfico corresponde a la función de distribución F(x) de una variable aleatoria con distribución normal estándar:



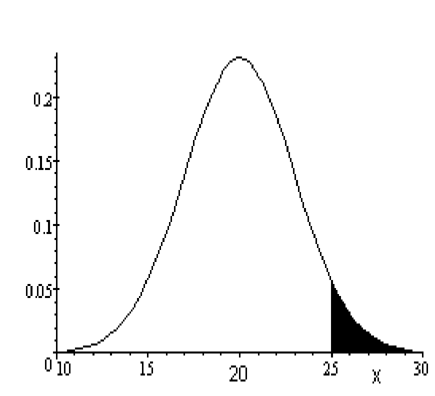
A partir de dicha figura resolver (aproximadamente):

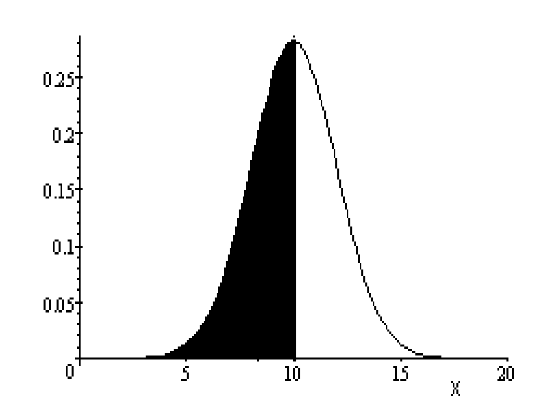
a)	*b)*
i) P(X<0)	**i)** F(x)=0.5
ii) P(X<4)	**ii)** 1-F(x)=0.1

6.

Encontrar el valor del área sombreada para las distribuciones normales con parámetros:

a) $\mu = 20$ $\sigma = 3$ ***b)*** $\mu = 10$ $\sigma = 2$





7.

Un especialista en ictiología tropical está interesado en estimar cuánto tiempo puede sobrevivir cierto tipo de pez en aguas con determinado porcentaje de toxicidad. Luego de una larga serie de experimentos llega a estimar que la vida media de este tipo de pez alcanza a los 90 días después de haber sido colocado en el agua venenosa, con una desviación estándar de 5 días. Es posible suponer que la distribución de los "días de sobrevida" es normal. ¿Cuál es la probabilidad de que un pez viva más de 120 días?.

8.

El peso de las raíces de alfalfa, que se puede considerar como índice de crecimiento, está distribuido normalmente con media 22 gr. y desviación estándar 10 gr. Hallar:

a) La probabilidad de que el peso de las raíces sea mayor a 23 gr.

b) La proporción de raíces con un peso inferior a los 10gr.

c) El peso tal que por debajo de él se encuentre el 50% de las raíces.

d) El peso tal que por encima de él se encuentren el 25% de las raíces (es decir el primer cuartil de la distribución).

9.

El peso de las langostas está distribuido normalmente con parámetros μ=6 gr. y σ=0.3 gr.

a) Hallar:
i) la proporción de langostas cuyo peso se encuentra entre 5.8 gr. y 6.4 gr.
ii) el porcentaje de langostas con un peso mayor a 6.7 gr.

b) Otra especie de langostas está distribuida normalmente con media μ= 5 gr. y desviación estándar 0.1 gr. Calcular las probabilidades solicitadas en los incisos anteriores bajo esta nueva distribución y compararlas con las anteriores.

c) ¿Cuál debiera ser el valor del peso medio de la población de tal modo que el 25% de los insectos tenga un peso superior a los 6 gr., suponiendo que la varianza de la población es la misma?

10.

Se observó durante un largo período que la cantidad semanal gastada, en mantenimiento y en reparaciones, en un tambo tiene aproximadamente una distribución normal con media $300 y una desviación estándar de $15. Si el presupuesto para la próxima semana es de $350.

a) ¿Cuál es la probabilidad de que los costos reales sean mayores que la cantidad presupuestada?.

b) Mediante una planificación del gasto es posible reducir la variabilidad del mismo pero no su media. ¿Cuánto debería ser la varianza de la cantidad semanal gastada de tal modo que la probabilidad de gastar más de 320$ sea prácticamente despreciable?.

11.

a) Si una población normal tiene media igual a 10 y varianza igual a 4.

a_1) ¿Qué porcentaje de las observaciones se encuentra entre 9 y 14?.

a_2) ¿Qué porcentaje de las observaciones se encuentra entre 13 y 15?.

b) Suponga que se tiene una distribución normal con media 0 y varianza 1.

b_1) ¿Entre qué valores se ubicará el 95% de las observaciones centrales?

b_2) ¿Entre qué valores se ubicará el 99% de las observaciones centrales?

12.

Sean las variables aleatorias X_1 , X_2 , X_3 , X_4 independientes N(0,1) hallar :

a) $P(\bar{X} < 1.4)$

b) $P(\bar{X} > 2.8)$

c) $P(-1 < \bar{X} < 1)$

d) $P(-2 < \bar{X} < 0.5)$

13.

Sean las variables aleatorias X_1 , X_2 , ... ,X_9 independientes N(3,2) hallar:

a) $P(\bar{X} < 2.5)$

b) $P(\bar{X} > 4.6)$

14.

En el estudio mencionado en el Ejercicio 1-f del Capítulo 1, suponga que se cuenta con la información de que la variable “porcentaje de festuca” sigue la distribución normal de media 50% con una desviación del 4%. Para un grupo de 25 vicuñas, ¿cuál es la probabilidad de que

a) el porcentaje medio de Festuca sea superior al 60%?

b) el porcentaje medio de Festuca se encuentre entre 45% y 55%.?

15.

Los huevos de las ponedoras Leghorn se embalan en cartones de 2 docenas cada uno. Los controles de calidad exigen que el peso medio por cartón sea por lo menos 40 gramos. Si el peso del cartón sigue una distribución normal de media 45 gr. y

desviación estándar 2 gr., ¿cuál es la probabilidad de que un cartón sea rechazado según las especificaciones del control de calidad?.

16.

El contenido de sales en el suelo tiene distribución normal con media 30.1 ppm y desviación estándar 7 ppm. Se seleccionan 9 muestras de suelo de la zona sur de la provincia de Córdoba. Calcular la probabilidad de que:

a) El contenido de sales del suelo sea superior a 22.5 ppm.

b) El contenido promedio de sales del suelo sea superior a 22.5 ppm.

c) El contenido promedio de sales del suelo esté entre 27.8 y 31.5 ppm.

6 Estimación Puntual y por Intervalos

Objetivos:

- Reconocer la utilidad de los intervalos de confianza como un método de estimación.
- Interpretar la información brindada por un intervalo de confianza.

6.1 Introducción

Como se estableció en los Capítulos 1 y 3 uno de los objetivos de la estadística es hacer inferencia con respecto a la población en base a la información contenida en una muestra. Según Lehmann (1983) "la inferencia es una adivinanza educada".

Las variables que se definen en los experimentos en general pueden ser asociadas a distintas distribuciones que dependen de parámetros, por ejemplo:

- La distribución Binomial está definida por una función de densidad puntual $f(x,n,p)$ que depende de los parámetros $\mathbf{n}$ y $\mathbf{p}$.
- La distribución Poisson está definida por una función de densidad puntual $f(x,\lambda)$ que depende del parámetro λ.
- La distribución Normal está definida por una función de densidad $f(x,\mu,\sigma)$ que depende de los parámetros μ y σ.

El objetivo de la mayoría de las investigaciones científicas es hacer *inferencia* con respecto a ciertos parámetros poblacionales, los cuales en general son desconocidos y por lo tanto el problema es obtener la mayor información posible acerca de ellos. Si fueran conocidos la función de densidad estaría totalmente determinada y no se debería procurar ninguna información sobre ellos. Entonces, utilizando la información contenida en una muestra aleatoria *única*, se intenta estimar el valor numérico del o de los parámetros desconocidos de interés.

Por ejemplo si la variable de interés tiene distribución Normal, entonces se puede estar interesado en estimar μ y/o σ, en tanto que si la variable de interés tiene distribución Binomial se puede estar interesado en estimar la proporción $\mathbf{p}$.

La *Inferencia Estadística* utiliza dos técnicas para lograr su objetivo: la *Estimación* (*puntual* y *por Intervalos*) y la *Prueba de Hipótesis*.

La estimación tiene muchas aplicaciones prácticas, por ejemplo:

* Un veterinario desea determinar la proporción de animales de una cierta raza afectados por una determinada enfermedad en la provincia de Buenos Aires.
* Un fabricante de lavadoras podría estar interesado en estimar la proporción de lavadoras que se descomponen antes de que termine el período de garantía de un año.
* Se desea estimar la cantidad promedio de mercurio que puede extraerse (mediante un proceso particular) de 1 onza de mineral proveniente de una localidad geográfica

particular.
* Para optimizar la atención en un supermercado se podría estar interesado en estimar el tiempo medio de espera en una caja registradora.
* Para conocer la precisión de un instrumento electrónico, se puede desear estimar la desviación estándar de la medición de dicho instrumento.

Retomando ahora el Problema 1.1, “Mediante una nueva dieta se espera que se incremente el peso en novillos de raza Charolais”. La variable aleatoria en estudio X: *Peso* tiene asociada una distribución de probabilidades que puede suponerse Normal. En esta etapa el objetivo es obtener información acerca del peso promedio de todos los novillos de raza Charolais alimentados con la nueva dieta, esto es $E(X)=\mu$ (μ es además uno de los parámetros de la distribución Normal).

Es natural pensar que para estimar el valor de la $E(X)=\mu$ se utilizará la información de la muestra, aunque evidentemente sobre la base de una muestra de tamaño $\mathbf{n}$ no se puede reconstruir la verdadera composición de la población en estudio. En otras palabras, a no ser que se inspeccione a cada animal de la población no se podrá conocer el verdadero valor de $E(X)=\mu$. Así surge la idea de buscar un valor aproximado para el parámetro. Por ejemplo para aproximar la media poblacional $E(X)=\mu$ se puede proponer a la media muestral $\overline{X}$. En este caso se está tratando a este estadístico como un *estimador*, término que será definido a continuación.

Definición 1: Un *estimador* es un estadístico que se utiliza para estimar un parámetro de la distribución.

6.2 Estimación Puntual

Para la estimación de un parámetro se podría utilizar solo un número; la intención es que dicho número esté cerca del verdadero valor del parámetro.

Si X es una variable aleatoria con distribución Normal cuyas características numéricas $E(X)=\mu$ y $Var(X)=\sigma^2$ son desconocidas y el interés es estimar dichas características, se pueden proponer como estimadores los estadísticos $\overline{X}$ y S^2, respectivamente.

Definición 2: Una *estimación puntual* de un parámetro es el valor que toma un estimador para una muestra particular.

Para el Problema 1.1 una estimación puntual del peso promedio de todos los novillos de raza Charolais con la nueva dieta es $\overline{X} = 413$ kg.

Para estimar el verdadero valor del parámetro, se puede elegir cualquier estadístico definido en el Capítulo 2. Lo que ocurre es que los estadísticos deben cumplir ciertas propiedades para ser considerados *buenos estimadores* y se eligen aquellos que cumplan con ellas. Ni éstas ni los métodos para generar buenos estimadores serán mencionados en este texto, pues están fuera del alcance del mismo (Mendenhall, W. et. al. - 1994).

Suponga que el problema es estimar la proporción de éxitos $\mathbf{p}$ de una variable aleatoria X con distribución Bernoulli. Para ello se extrae una muestra aleatoria de tamaño $\mathbf{n}$ de dicha distribución, denotada por $(X_1, X_2, ..., X_n)$. Si se define $Y = \sum_{i=1}^{n} X_i$ como el número de éxitos en

la muestra, un buen estimador puntual del parámetro **p** resulta $\hat{p} = \frac{Y}{n}$.

Dado que los estadísticos $\overline{X}$ y S^2 verifican las propiedades requeridas para ser un buen estimador, son utilizados generalmente como estimadores de la $E(X)$ y la $Var(X)$ respectivamente.

Como la estimación raramente coincide con el parámetro y no se puede cuantificar la diferencia entre ellos, surge un segundo procedimiento de estimación.

6.3 Estimación por Intervalo

La *Estimación por Intervalo* es un método que permite estimar un parámetro generando dos números a partir de la estimación puntual del mismo. Éstos números son denominados límite inferior y límite superior de un intervalo que se espera que incluya al verdadero valor del parámetro. Este intervalo se denomina *intervalo de confianza*.

Idealmente sería conveniente que el intervalo tuviera las siguientes propiedades: que contenga al verdadero valor del parámetro y que sea relativamente de longitud pequeña.

Los límites del intervalo de confianza son funciones de estimadores puntuales y por tanto son variables aleatorias. Entonces, la idea es construir un intervalo aleatorio que, con *cierta certeza* (confianza), cubra al verdadero valor a ser estimado.

Definición 3: El *nivel de confianza*, $1-\alpha$, es la probabilidad de que el intervalo aleatorio contenga al verdadero valor del parámetro.

Desde el punto de vista práctico, el nivel de confianza indica la fracción de veces, en un muestreo repetitivo, que los intervalos contendrán al parámetro de interés. Si el nivel de confianza asociado al intervalo fuera alto, entonces se estaría altamente confiado de que un intervalo de confianza particular, construido a partir de una sola muestra, contenga al parámetro de interés. Por ello se utilizan habitualmente niveles de confianza altos (0.90, 0.95 y 0.99).

6.3.1 Intervalo de Confianza para la media de una Distribución Normal

Suponga que se desea realizar una estimación por intervalo del parámetro μ, de una distribución normal.

Dado que no es seguro que cualquier intervalo (I,S) contenga a μ, puede afirmarse que

$$P\big(\ I(X_1,X_2,...,X_n)) < \mu < S(X_1,X_2,...,X_n)\ \big) = 1 - \alpha$$

Para determinar los límites del intervalo se procede de la siguiente forma:

1. Sobre la base de una muestra de tamaño **n**, se determina el estimador puntual $\overline{X}$ del parámetro μ. De la estandarización de la variable aleatoria $\overline{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$ resulta la

variable aleatoria $Z = \frac{(\overline{X} - \mu)}{\sigma / \sqrt{n}}$ que tiene distribución $N(0,1)$.

2. Como Z involucra al parámetro a estimar (μ) y a su estimador puntual ($\overline{X}$), para encontrar el intervalo de confianza se plantea la igualdad $P(-a < Z < a) = 1-\alpha$. Fijando el nivel de confianza $1-\alpha$ se obtiene el valor de abscisa **a** de la Tabla C del Apéndice. Gráficamente

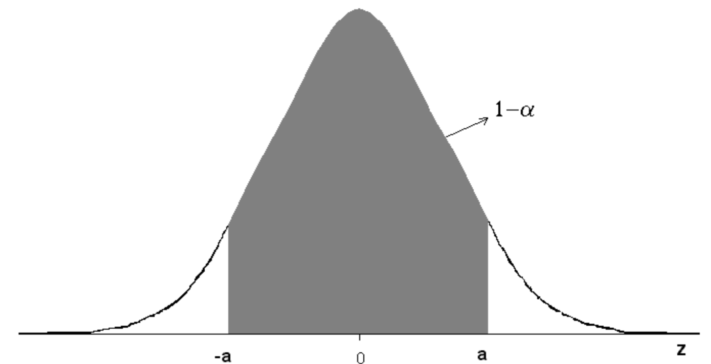


Gráfico 1: Valores de abscisa correspondientes a la probabilidad $1-\alpha$

3. A partir de la expresión $P(-a < Z < a) = 1-\alpha$ reemplazando la variable Z se tiene

$$P\left(-a < \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < a \right) = 1 - \alpha$$

Dado que se está interesado en determinar un intervalo para μ, a partir de la expresión anterior se trabaja algebraicamente de la siguiente manera:

$$P\left(-a \frac{\sigma}{\sqrt{n}} < (\overline{X} - \mu) < a \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

$$P\left(-a \frac{\sigma}{\sqrt{n}} - \overline{X} < -\mu < a \frac{\sigma}{\sqrt{n}} - \overline{X} \right) = 1 - \alpha$$

$$P\left(\overline{X} - a \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + a \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha \qquad (6.1)$$

De donde se pueden obtener las siguientes expresiones explícitas de los límites Inferior (I) y Superior (S):

$$I = \overline{X} - a \frac{\sigma}{\sqrt{n}} \qquad \text{y} \qquad S = \overline{X} + a \frac{\sigma}{\sqrt{n}}$$

Es importante notar que los límites del intervalo son variables aleatorias dado que, en este caso, $\overline{X}$ es una variable aleatoria. Luego se ha construido un *Intervalo Aleatorio* con nivel $1-\alpha$ de confianza para la media μ de una población Normal con desviación σ conocida.

En cuanto al significado del Nivel de Confianza, si éste fuera 0.90, se espera que de cada *100 intervalos de confianza* construidos a partir de 100 muestras de tamaño fijo **n**, aproximadamente *90 de ellos contengan el verdadero valor de μ y 10 intervalos no.*

Gráficamente se tiene

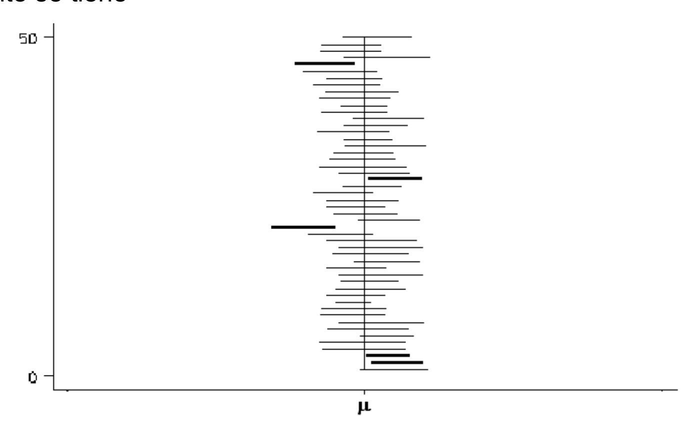


Gráfico 2: 50 Intervalos de confianza del 90% para μ ,
con muestras de tamaño fijo

Se puede observar claramente que μ es un valor fijo (desconocido) y que los intervalos son aleatorios y “tratan” de cubrir a μ.

Retomando el Problema 1.1, suponiendo que el peso de los novillos de la raza Charolais a los que se les suministró la nueva dieta tiene un desvío estándar de 4.18 kg. resulta que $X \sim N(\mu, 4.18)$. Se seleccionan 16 novillos de la raza en estudio y se les aplica la nueva dieta, obteniendo un peso promedio de 413 kg. En base a esta información se construye el intervalo de confianza para el peso medio μ de todos los novillos a los que se les podría aplicar la nueva dieta.

Para este caso se tienen los siguientes datos: $\sigma=4.18$, $n=16$ y $\overline{X} = 413$. Si se fija un nivel de confianza $1-\alpha=0.99$, de la expresión $P(-a< Z < a) = 0.99$ se obtiene el valor de abscisa $a=2.57$ (Tabla C).

Substituyendo a, σ y **n** en la expresión (6.1) se obtiene

$$P\left(\overline{X} - 2.57 \bullet \frac{4.18}{4} < \mu < \overline{X} + 2.57 \bullet \frac{4.18}{4}\right) = 0.99$$

Cuando se reemplaza la media muestral por el valor particular, la expresión anterior deja de contener la variable aleatoria $\overline{X}$ y por lo tanto no tiene sentido calcular una probabilidad. Entonces se reemplaza la **P** de probabilidad por la **C** de confianza, o sea

$$C\left(413 - 2.57 \bullet \frac{4.18}{4} < \mu < 413 + 2.57 \bullet \frac{4.18}{4}\right) = 0.99$$

$$C(410.31 < \mu < 415.69) = 0.99$$

Esta expresión debe ser interpretada de la siguiente manera “existe una confianza del 99% de que el intervalo (410.31 , 415.69) cubra al peso promedio de todos los novillos Charolais con la nueva dieta”.

Dado que una de las características que se espera de un intervalo de confianza es que tenga longitud pequeña, es interesante ver cómo manejar esta longitud.

La precisión de un intervalo está en función de su longitud, en el sentido de que un intervalo de longitud grande será considerado de poca precisión. Hay dos formas de construir intervalos de confianza más precisos:

- Disminuyendo el nivel de confianza $1-\alpha$.
- Aumentando el tamaño de muestra $\mathbf{n}$.

Para entender estas afirmaciones basta observar la expresión (6.1).

En la mayoría de los casos reales la desviación estándar poblacional será desconocida, luego el intervalo presentado no es el más utilizado. En ese caso se usa la varianza muestral S^2 para estimar a σ^2 y el intervalo de confianza para la media μ, con un nivel de confianza $(1-\alpha)$, resulta

$$P\left(\overline{X} - a \bullet \frac{S}{\sqrt{n}} < \mu < \overline{X} + a \bullet \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

donde la variable $Z = \dfrac{\overline{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$ y por lo tanto la abscisa $a = t_{n-1, 1-\alpha/2}$ se obtiene de la Tabla D del Apéndice.

6.3.2 Intervalo de Confianza para varianza de una Población Normal

Existen situaciones en la práctica donde el parámetro de interés es la varianza poblacional; por ejemplo cuando se desea conocer la variabilidad en instrumentos de medición si la variable en estudio tiene distribución Normal el problema se reduce a generar un Intervalo de confianza para la varianza de una distribución Normal (σ^2).

El estimador puntual de la varianza poblacional es la varianza muestral S^2, para la que se verifica que $\dfrac{(n-1)\, S^2}{\sigma^2} \sim \chi^2_{n-1}$ (ver Sección 5.5.3). Fijando un nivel de confianza $1-\alpha$ y trabajando algebraicamente a partir de la siguiente igualdad

$$P\left(a < \frac{(n-1)S^2}{\sigma^2} < b\right) = 1 - \alpha$$

se obtiene el Intervalo de confianza para la varianza de una población normal

$$P\left(\frac{1}{b}(n-1)S^2 < \sigma^2 < \frac{1}{a}(n-1)S^2\right) = 1 - \alpha$$

de donde resultan los límites $I = \dfrac{1}{b}(n-1)S^2$ y $S = \dfrac{1}{a}(n-1)S^2$, con $a = \chi^2_{n-1, 1-\alpha/2}$ y $b = \chi^2_{n-1, \alpha/2}$ (Tabla E del Apéndice).

6.3.3 Intervalo de Confianza para la proporción de una Distribución Binomial

Suponga que se desea construir un intervalo de confianza para la proporción de éxito $\mathbf{p}$,

de una distribución Binomial. Para ello se tomará el siguiente

Problema 6.1: El dueño de una cabaña desea comprar animales vacunos de una cierta zona y antes de realizar la compra desea estimar la proporción de animales afectados por aftosa en dicha zona. Para dar una respuesta al dueño de la cabaña se seleccionaron al azar 20 animales vacunos de la zona registrándose si estaban o no afectados con aftosa.

La variable que se observa es X: "Número de animales vacunos afectados entre los 20 seleccionados", la cual tiene distribución $B(20,p)$.

Para construir el intervalo de confianza es necesario determinar el estimador puntual del parámetro, su esperanza y su varianza, para lo cual se procede de la siguiente manera:

1. Si en las $\mathbf{n}$ repeticiones del Ensayo de Bernoulli hubo X éxitos, el estimador puntual para la proporción de éxitos es $\hat{p} = \dfrac{X}{n}$.

2. Como el estimador $\hat{p}$ es una variable aleatoria se pueden determinar las características numéricas tales como
$E(\hat{p}) = E\left(\dfrac{X}{n}\right) = \dfrac{1}{n}E(X) = p$ y $Var(\hat{p}) = Var\left(\dfrac{X}{n}\right) = \dfrac{1}{n^2}Var(X) = \dfrac{p \bullet (1-p)}{n}$,

Estos resultados se obtienen utilizando las propiedades E_2 y V_2 del Capítulo 4.

3. Por la relación entre la distribución Normal y la distribución Binomial, válida para valores de $\mathbf{n}$" razonablemente grande" (dependiendo de lo cercano que esté $\mathbf{p}$ de $1/2$)

$$\hat{p} \sim N\left(p, \sqrt{\frac{p \bullet (1-p)}{n}}\right) \text{ entonces } Z = \frac{\hat{p} - p}{\sqrt{p \bullet (1-p)/n}} \sim N(0,1).$$

4. La variable Z no puede utilizarse para encontrar los límites del intervalo dado que el desvío de $\hat{p}$ ($\sqrt{p \bullet (1-p)/n}$) depende del parámetro desconocido $\mathbf{p}$. Para resolver este problema se recurre al estimador puntual de $\mathbf{p}$ resultando el desvío estimado $\sqrt{\hat{p} \bullet (1-\hat{p})/n}$. A partir de esto surge la variable $Z' = \dfrac{\hat{p} - p}{\sqrt{\hat{p} \bullet (1-\hat{p})/n}}$ que tiene distribución aproximadamente Normal y puede utilizarse para construirse el intervalo deseado.

Entonces el intervalo de confianza para el parámetro $\mathbf{p}$ de nivel $(1-\alpha)$ está dado por

$$P\left(\hat{p} - a\sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + a\sqrt{\hat{p}(1-\hat{p})/n}\right) = 1 - \alpha$$

donde $\mathbf{a}$ se obtiene de la Tabla C del Apéndice.

Retomando el Ejemplo 4, si de 20 animales resultaron 15 afectados por aftosa y se fija un nivel de confianza del 95%, el estimador puntual de $\mathbf{p}$ resulta $\hat{p} = \dfrac{X}{n} = \dfrac{15}{20} = 0.75$ y por lo tanto el intervalo de confianza es

$$C\left(0.75 - 1.96 \bullet \sqrt{\frac{0.75 \bullet 0.25}{12}} < p < 0.75 + 1.96 \bullet \sqrt{\frac{0.75 \bullet 0.25}{12}}\right) = 0.95$$

$$C(0.505 < p < 0.995) = 0.95$$

Esto se interpreta como “hay una confianza del 95% que la proporción de animales afectados por aftosa es un valor que está entre el 0.505 y el 0.995”.

Ejercicios de Aplicación

1.

Se desea estudiar el efecto sobre el aumento de peso de una dieta rica en vitamina A, en ratas de la especie Calomys Musculynus desde el nacimiento hasta la edad de tres meses. Estudios anteriores han determinado que la variable aumento de peso puede ajustarse con una distribución normal, de varianza 9 gr^2. Para estimar el aumento de peso promedio, se suministró la dieta a 16 de estas ratas obteniendo una media de 61 gr.

a) ¿Cuál es la variable que permitirá estudiar el efecto de la nueva dieta? ¿A qué tipo corresponde esta variable y cuál es su distribución de probabilidades?.

b) ¿Qué se desea estimar ? ¿Ese valor, es de la muestra o de la población? ¿Cuál es su relación con la variable?.

c) ¿Con que información poblacional y muestral se cuenta?.

d) Para el problema de la estimación por intervalos ¿Qué variable se construye y cuál es su distribución?.

e) Estimar el parámetro en cuestión con una confianza del 95% e indicar las conclusiones.

2.

Un dermatólogo que investiga cierto tipo de cáncer de piel, desea estimar el tiempo promedio hasta la desaparición del mismo con un nuevo fármaco (suponga que el tiempo se distribuye normalmente). Para ello induce este cáncer en 25 ratas y las trata, obteniéndose un promedio de 132 hs. para la desaparición de la enfermedad con un desvío estándar de 101 hs.

a) ¿Cuál es la variable que permitirá estudiar el comportamiento del nuevo medicamento? ¿A qué tipo corresponde?.

b) ¿Qué se desea estimar ? ¿Ese valor, es de la muestra o de la población? ¿Cuál es su relación con la variable?.

c) ¿ Cuál es la información poblacional y cuál es la brindada por el experimento?.

d) Efectuar la estimación del parámetro de interés, con una confianza del 99% e indicar las conclusiones.

3.

Se desea estimar con una confianza del 99%, el contenido promedio de alquitrán de cierta marca de cigarrillos (suponga que el contenido de alquitrán se distribuye normalmente). Para ello se selecciona una muestra de 36 cigarrillos, obteniéndose una media de 17,2 mg. y una desviación estándar de 2 mg.

4.

Un grupo de investigación desea estimar el porcentaje medio de sacarosa en la caña de azúcar luego de habérsele aplicado un producto compuesto con sales de molibdeno y otros metales que tienden a inhibir procesos enzimáticos. (suponga que el porcentaje se distribuye normalmente). Para ello fueron seleccionadas 7 plantas de caña de azúcar a las cuales de les aplicó el producto, obteniéndose una media de 0,84% de sacarosa y un desvío de 0,18% de sacarosa.

a) ¿Cuál es la variable en estudio? ¿Qué representan los parámetros de la distribución de la variable en este problema ?

b) Estimar el porcentaje medio de sacarosa con un 90% de confianza.

c) ¿Podría mejorar la precisión del intervalo de confianza para el porcentaje medio de sacarosa? ¿Cómo lo realizaría?

5.

Los límites de confianza del 95% para la media de una poblacional son 20 y 30. ¿Cuál de las siguientes afirmaciones es correcta?

a) De cien medias muestrales extraídas al azar de esta población cerca de 95 de ellas estarán entre 20 y 30.

b) De cien medias poblacionales extraídas al azar de esta población cerca de 95 estarán entre 20 y 30.

c) Hay una confianza del 95% que los límites 20 y 30 cubran a la verdadera media poblacional.

6.

A continuación se presentan los promedios y desvíos correspondientes a tres muestras de tamaño 16 provenientes de una población con distribución Normal de media μ=80:

Muestra	Media	Desvío
1º	77.76	8.45
2º	74.46	5.27
3º	78.58	7.29

a) A partir de los datos muestrales construir, para cada muestra, un intervalo de confianza del 95% para la media de la distribución.

b) ¿Contienen los intervalos al parámetro en cuestión?

c) Explique los resultados de lo ocurrido en b) a partir del significado de la confianza.

7.

Se desea estimar, con un nivel del 90%, la proporción de personas que tienen sangre de tipo A positivo en una cierta ciudad. Para ello se tomó una muestra aleatoria de 400 personas encontrándose 125 con sangre de tipo A positivo.

8.

En el contexto del Ejercicio 10 del Capítulo 5, suponga que se desconoce el valor del gasto medio por semana, y que sólo se sabe de la variable gasto semanal que su distribución es normal. Si se elige una muestra de 30 semanas ¿qué valor debería asumir la varianza de tal modo que, con un 99% de confianza, la media del gasto semanal se encuentre entre $290 y los $330?.

7 Prueba de Hipótesis

Objetivos:

- Analizar un problema dado y plantear las hipótesis correspondientes.
- Interpretar la información obtenida a través de una prueba de hipótesis.
- Distinguir entre Intervalo de Confianza y Prueba de Hipótesis en cuanto a su utilidad.

7.1 Introducción

La Inferencia estadística brinda métodos que permiten, a través de una Muestra, obtener información acerca de alguna característica de la Población de la cual fue extraída.

En el Capítulo 6 se estudiaron dos métodos de estimación denominados Estimación Puntual y Estimación por Intervalos. A continuación se presenta otro método de inferencia llamado Prueba de Hipótesis (o Test de Hipótesis).

Definición 1: Una *Prueba o Contraste de Hipótesis* es un procedimiento mediante el cual se investiga la verdad o falsedad de una hipótesis contrastada.

7.2 Prueba de hipótesis para la media de una Distribución Normal

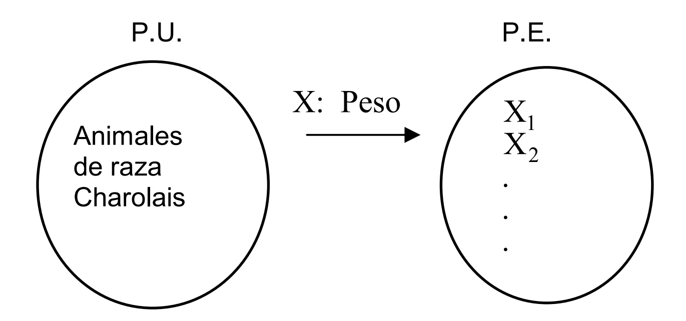
En base al Problema 1.1 se plantea la siguiente situación: "Mediante una nueva dieta se espera que se incremente el peso medio de novillos de la raza Charolais de determinado peso inicial".

En otras palabras, lo que se desea es confirmar si la *nueva dieta es mejor* que la dieta usual. Surge entonces la necesidad de realizar una comparación, en realidad se desea probar, por ejemplo, si el peso medio obtenido con la nueva dieta es mayor que el peso medio obtenido con la dieta utilizada anteriormente. Para resolver este problema se utilizará una prueba de hipótesis, para lo que es necesario en primer lugar traducir lo anterior en términos estadísticos.

Se comienza por identificar para este caso los elementos necesarios para realizar el análisis estadístico.

- Unidad experimental: Un novillo de raza Charolais, con un cierto peso inicial.
- Población de Unidades: Todos los novillos de raza Charolais de cierto peso inicial.
- Variable en estudio: Peso
- Población Estadística: El peso de todos los novillos de cierto peso inicial raza Charolais.

La situación anterior puede ser esquematizada como sigue



Las características numéricas de una población estadística son $E(X)$ y $Var(X)$, que en este caso indican el peso promedio y la varianza del peso de todos los animales sujetos a la nueva dieta, ambos desconocidos.

Por resultados de experiencias previas se sabe que el *peso medio* de todos animales con la dieta que se venía utilizando, es de 390 kg., valor que se puede denotar como μ_0; luego lo que se desea es comparar este valor con el *peso medio* de todos los animales con la nueva dieta, lo que se puede expresar de la siguiente forma

I)
H_0: El peso medio de todos los animales con la nueva dieta es igual al peso medio de todos los animales con la dieta usada habitualmente.

II)
H_1: El peso medio de todos los animales con la nueva dieta es superior al peso medio de todos los animales con la dieta usada habitualmente.

Las situaciones I) y II) pueden ser escritas en lenguaje estadístico, como se indica a continuación

$$H_0: \mu = \mu_0 \qquad\qquad H_1: \mu > \mu_0$$

donde

μ : indica el peso medio de todos los novillos alimentados con la nueva dieta.

μ_0: indica el peso medio de todos los novillos alimentados con la dieta usada habitualmente.

La variable peso puede asumir dos distribuciones normales diferentes,

$$X: \text{Peso} \begin{cases} \rightarrow N(\mu_0,\sigma_0) \\ \rightarrow N(\mu,\sigma) \end{cases}$$

Considerando que la nueva dieta puede modificar la media pero no el desvío estándar, se supone $\sigma = \sigma_0$. Las dos distribuciones que la variable X asume pueden visualizar gráficamente de la siguiente manera

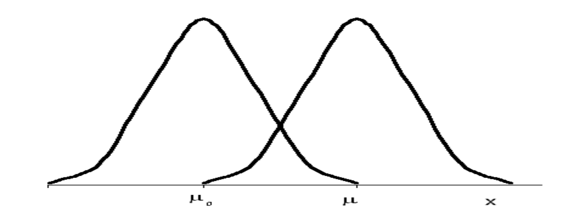


Gráfico 1: Funciones de densidad correspondientes a $X \sim N(\mu_0,\sigma_0)$ y $X \sim N(\mu,\sigma_0)$

Para este problema lo que se desea probar es:

H_0: $\mu = 390$ kg. H_1: $\mu > 390$ kg.

En general se puede dar la siguiente

Definición 2: La *Hipótesis Nula* (H_0) es la hipótesis estadística cuya verdad o falsedad se va a establecer estadísticamente usando los datos muestrales como evidencia. La *Hipótesis alternativa* (H_1) es cualquier otra hipótesis distinta de H_0.

En la prueba para la media de una población normal H_0:μ=μ_0. Mientras que H_1 puede plantearse de una (y sólo una) de las siguientes formas

a) H_1:$\mu > \mu_0$ b) H_1:$\mu < \mu_0$ c) H_1: $\mu \neq \mu_0$

Cuando la hipótesis alternativa es de la forma como la indicada en a) o b) la prueba es *unilateral* o de una cola y cuando la hipótesis alternativa es de la forma coma la indicada en c) se dice que la prueba es *bilateral* o de dos colas.

Para tomar una decisión sobre las hipótesis planteadas, se debe utilizar la información muestral. Para ello se realiza un experimento aleatorio (generando una muestra aleatoria de tamaño $\mathbf{n}$), que para el problema planteado consiste en elegir aleatoriamente 16 animales de raza Charolais, con cierto peso inicial, a los que se les suministra la nueva dieta. Después de un cierto tiempo (prefijado de antemano) se les mide el peso, generando así la muestra estadística, de la cual se obtuvo $\overline{X} = 413$ kg. y $S = 5.15$ kg.

Teniendo en cuenta que $\overline{X}$ es un *buen estimador* de $E(X) = \mu$, parece razonable que se base la decisión de rechazar o no H_0 en el valor de $\overline{X}$. Puesto que se está interesado en distinguir entre μ=390 kg. y $\mu > 390$ kg. se debería rechazar H_0 cuando ($\overline{X} - \mu_0$) sea "muy grande", o sea se rechazaría H_0 cuando $\left(\overline{X} - 390\right)$ sea mayor que una *cierta constante*; el problema ahora es determinar dicha constante.

Para determinar dicha constante además de la diferencia $\left(\overline{X} - 390\right)$ se debe tener en cuenta la variabilidad de la población. Toda la información para determinar la constante en cuestión puede ser resumida en una nueva variable aleatoria, que recibe el nombre de estadístico de contraste y es denotado por ε. Éste debe contener la información de la muestra y una distribución de probabilidades conocida.

Como ya se vio en el capítulo anterior, al construir un intervalo de confianza para la media de una distribución Normal es necesario saber si se conoce o no la varianza de la población en estudio, ya que la distribución de la variable aleatoria construida depende de ello. Para una prueba de hipótesis la variable aleatoria construida es el estadístico de contraste y las dos situaciones se presentan en las secciones siguientes.

Definición 3: El *Estadístico de Contraste* (ε) es una variable aleatoria (está en función de valores muestrales) cuya distribución de probabilidades depende de la validez de la hipótesis nula.

Cuando se realiza la prueba para la media de una población normal el estadístico ε, depende siempre de $\overline{X}$.

7.2.1 Con varianza conocida

Por lo expresado anteriormente, el estadístico de contraste

$$\varepsilon = \frac{\overline{X} - \mu_o}{\sigma / \sqrt{n}}$$

es una variable aleatoria (por ser una función de la variable aleatoria $\overline{X}$),que tiene asociada una distribución de probabilidades que se estudiará a continuación.

Dado que la variable en estudio X tiene distribución Normal con parámetros μ y σ, esto es, $X \sim N(\mu,\sigma)$ y si se tiene una muestra aleatoria de tamaño **n**, la variable $\overline{X} \sim N(\mu, \sigma / \sqrt{n})$ y por tanto $Z = \frac{(\overline{X} - \mu)}{\sigma / \sqrt{n}} \sim N(0,1)$.

Entonces

- si vale H_0 (o sea cuando $\mu=\mu_0$), $\varepsilon \sim N(0,1)$;
- si no vale H_0 (o sea cuando $\mu > \mu_0$), $\varepsilon \sim N(c,1)$, donde $c = \frac{\mu - \mu_o}{\sigma / \sqrt{n}}$.

Para determinar el valor de $\mathbf{c}$, se necesita cuantificar la diferencia $(\mu-\mu_0)$; comúnmente se la considera en términos de σ ($\mu-\mu_0=\sigma$, $\mu-\mu_0=2\sigma$ o $\mu-\mu_0=3\sigma$). El valor de esta diferencia siempre debe ser sugerido por el investigador.

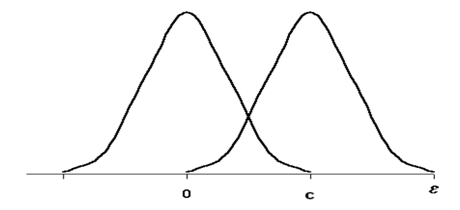


Gráfico 2: Funciones de densidad correspondientes
a $\varepsilon \sim N(0,1)$ y $\varepsilon \sim N(c,1)$

Por otro lado se debe tener en cuenta que al tomar la decisión de *rechazar* o de *no rechazar* la hipótesis nula H_0, no se hace con un 100% de seguridad, pues se está trabajando con la información de una muestra para obtener conclusiones con respecto a toda una población. Por lo tanto la decisión puede ser errónea, esto es, se puede rechazar H_0 cuando en realidad vale o bien se puede no rechazar H_0 cuando ésta no vale.

Se puede decir entonces que cuando se realiza una prueba de hipótesis pueden ocurrir las situaciones siguientes

	Vale H_0	**No vale** H_0
Rechazar H_0	***Error de tipo I***	***Decisión Correcta***
No Rechazar H_0	***Decisión Correcta***	***Error de tipo II***

Por lo tanto

Definición 4: El *Error de Tipo I* se comete cuando se rechaza H_0 y ésta en realidad vale. El *Error de Tipo II* se comete cuando no se rechaza H_0 y ésta en realidad no vale.

Así cualquiera de las dos decisiones que se pueden tomar, pueden ser equivocadas y es importante medir el riesgo de tomar una decisión errónea. Este riesgo se mide a través de una probabilidad, que por ser una probabilidad de error debe tomar valores pequeños.

Definición 5: El *nivel de significación de la prueba* (α) es la probabilidad de cometer error de tipo I, es decir α =P(cometer error tipo I)=P($\varepsilon \in Z$ / vale H_0). Gráficamente es el área bajo la curva sobre la zona de rechazo.

Como se dijo, la idea es determinar una constante a partir de la cual se pueda decir que la nueva dieta produjo mayor peso medio que la dieta habitual. Esta constante, denotada usualmente **a**, es el valor crítico y determina un intervalo llamado zona de rechazo (Z) .

Definición 6: La *Zona de rechazo* (Z) es el conjunto de valores del estadístico de contraste que lleva a descartar la H_0. La constante **a** asociada a Z es el *valor crítico.*

Para el Problema, la zona de rechazo es

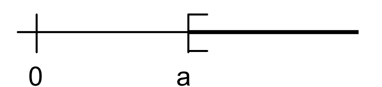


Gráfico 3: Zona de Rechazo

Para determinar el valor de **a** se debe tener en cuenta H_1, la distribución del estadístico de contraste bajo H_0 y el valor de α. Para este caso particular se toma $\alpha=0.01$ como nivel de significación de la prueba. En el siguiente gráfico se pueden observar el nivel de significación y su correspondiente zona de rechazo.

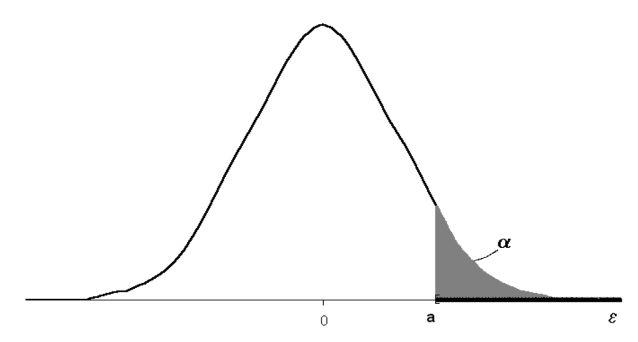


Gráfico 4: Zona de rechazo y nivel de significación

Luego para encontrar el valor numérico de la constante **a** se plantea la siguiente expresión $P(\varepsilon \sim N(0,1) \geq a)=0.01$, de donde se obtiene $a=2.32$.(Tabla C del Apéndice).

El intervalo $[a,+\infty)=[2.32,+\infty)$ es entonces la Zona de Rechazo y la constante $a=2.32$ es el Valor Crítico de la prueba.

Notar que los valores que puede asumir el estadístico de contraste son números reales (gráficamente se ubican en el eje de las abscisas o eje x).

Gráficamente:

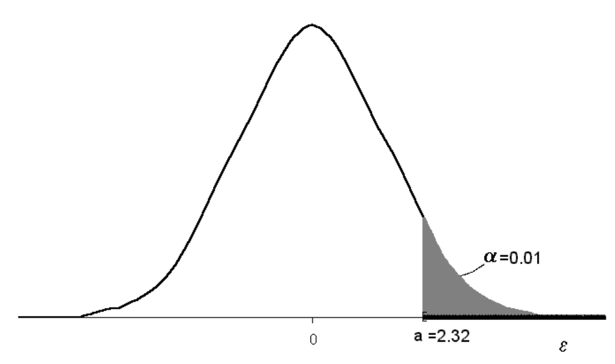


Gráfico 5: Zona de rechazo y nivel de significación
para el Problema 1.1

Para tomar la decisión se debe calcular el valor numérico del estadístico, que se denota por ε_c. Para determinar dicho valor es necesario usar la desviación estándar de la población en estudio, que en este caso se supone conocida e igual a 4.18. Luego el valor numérico del estadístico resulta

$$\varepsilon_c = \frac{413-390}{4.18/\sqrt{16}} = 22.01$$

A partir de la zona de rechazo y del valor numérico del estadístico se establece una regla para tomar una decisión:

- Si $\varepsilon_c \in Z$ se rechaza H_0.
- Si $\varepsilon_c \notin Z$ no se rechaza H_0.

Para este problema, el valor numérico del estadístico es $\varepsilon_c=22.01$, que pertenece a la zona de rechazo, luego se rechaza la hipótesis nula con una probabilidad de cometer error de tipo I de 0.01.

Conclusión: *El peso medio de los animales con la nueva dieta es mayor que el peso medio de los animales con la dieta usada habitualmente.*

En la secuencia presentada anteriormente se fija el valor de α antes de realizar el experimento. No siempre se puede o se quiere especificar este valor, por ejemplo porque la persona que toma la decisión no es quien realiza la investigación. Por otra parte la elección de α es arbitraria y puede suceder que con el mismo conjunto de datos se llegue a conclusiones diferentes. Una alternativa para evitar esta ambigüedad es determinar en primer lugar el valor del ε_c (después de realizado el experimento) y luego calcular una probabilidad que depende de éste, llamada valor **p** de la prueba.

Definición 7: El *valor* **p** *de la prueba* es la mínima probabilidad de error de Tipo I que lleva a rechazar la Hipótesis nula.

Cuando la hipótesis alternativa es $H_1\!:\mu > \mu_0$, el valor **p** se calcula como

$$p = P(\, \varepsilon \geq \varepsilon_c \,)$$

que se manifiesta en el siguiente gráfico

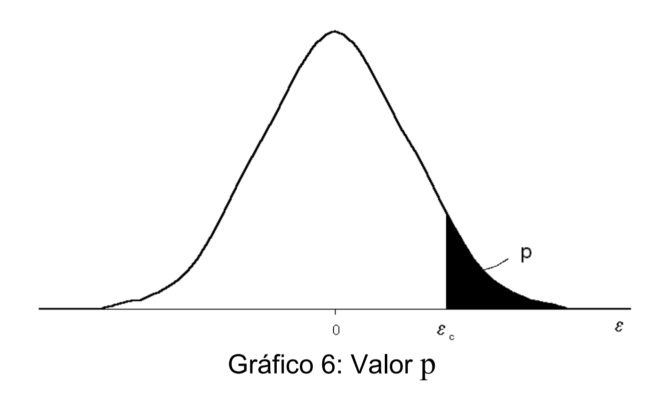


Gráfico 6: Valor p

Para tomar la decisión en base al valor **p** se debe fijar la máxima probabilidad de error de tipo I que se está dispuesto a cometer, y compararla con el valor **p**. Si el valor **p** es menor que esta probabilidad se rechaza la hipótesis nula con probabilidad exacta **p** de cometer error de tipo I.

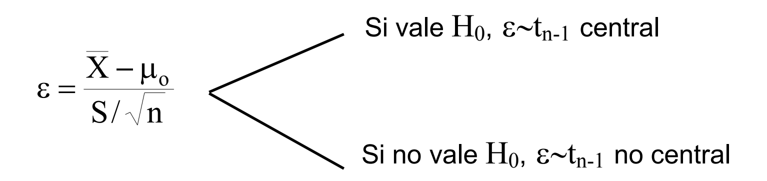
Notar que el valor **p** de la prueba depende de la hipótesis alternativa, de la distribución del estadístico y de la información brindada por la muestra.

Para el Problema 1.1 como $H_1: \mu > 390$ kg., $\varepsilon \sim N(0,1)$ si vale H_0 y $\varepsilon_c = 22.01$, el valor **p** es **p** $= P(\varepsilon \geq 22.01) \cong 0$ por lo cual se rechaza H_0 con probabilidad prácticamente nula de cometer error de tipo I.

7.2.2 Con varianza desconocida

En la mayoría de los casos el investigador no tiene ninguna sospecha sobre el valor de σ. En este caso la prueba de hipótesis para la media de una distribución Normal se construye de manera similar a lo desarrollado en la sección anterior. La diferencia está en que se usa **S** como estimador de σ, lo que determina un cambio en la expresión y en la distribución del *estadístico* ε.

Así, en este caso se tiene que



donde t_{n-1} indica la distribución t - Student con $n-1$ grados de libertad (g.l.) centrada o no alrededor de cero.

Notar que, al cambiar la distribución del estadístico, el valor crítico se obtiene de la Tabla t de Student.

En el caso que el investigador tenga alguna sospecha sobre el valor de σ, se puede realizar una prueba de hipótesis como la que se desarrolla en la Sección 7.3. En dicha prueba, si se rechaza H_0 el estadístico que corresponde es el de la presente sección; si no se rechaza H_0 se puede suponer $\sigma = \sigma_0$ y usar el estadístico presentado en la sección anterior.

7.2.3 Algunas consideraciones importantes

Así como el riesgo de cometer error de tipo I se mide con una probabilidad (nivel de significación), el riesgo de cometer error de tipo II también se mide con una probabilidad:

$1-\beta$=P(cometer error de Tipo II) = P(no rechazar H_0 dado que no vale H_0).

Por otra parte la probabilidad de no cometer error de tipo II se llama *Potencia de la prueba* y es

β=P(no cometer error de Tipo II) = P(rechazar H_0 dado que no vale H_0)

La potencia de la prueba indica la capacidad o "poder" de la prueba para rechazar correctamente una H_0 falsa.

Gráficamente las probabilidades de error de tipo I y II y la potencia de la prueba son:

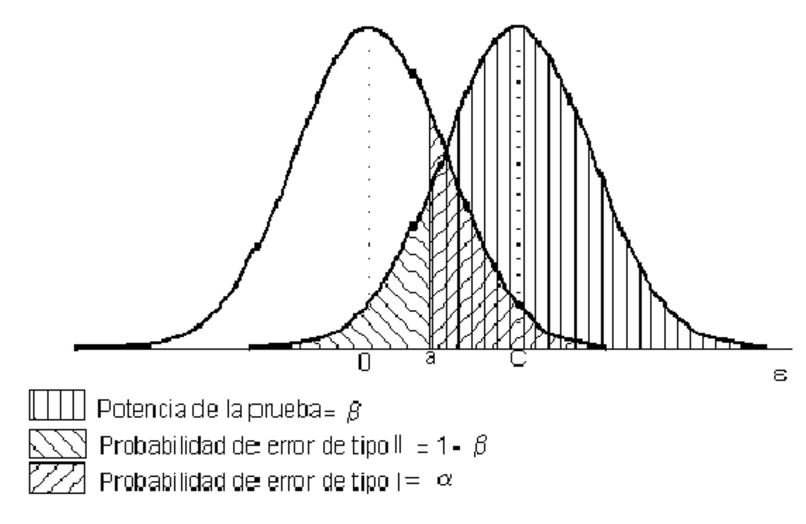


Gráfico 7: Probabilidades de error de tipo I y II y
potencia de la prueba

Cuando el investigador decide fijar el valor de α debe analizar qué tipo de error es más grave cometer (si el Error de Tipo I o el de Tipo II) . Si es más grave el Error de Tipo I deberá tomar un valor pequeño de α, teniendo en cuenta que de esa manera se aumenta la probabilidad de cometer el Error de Tipo II. Y si es más grave cometer el Error de Tipo II deberá fijar un valor de α más grande, porque de esa manera se asegura que disminuya la probabilidad de cometer Error de Tipo II (observar Gráfico 7).

La posibilidad de cometer un error (de Tipo I o II) siempre está presente, pues se trata de obtener alguna información sobre la población a través de una muestra (aunque ésta sea representativa de la población).

La forma de la zona de rechazo depende de la *hipótesis alternativa*

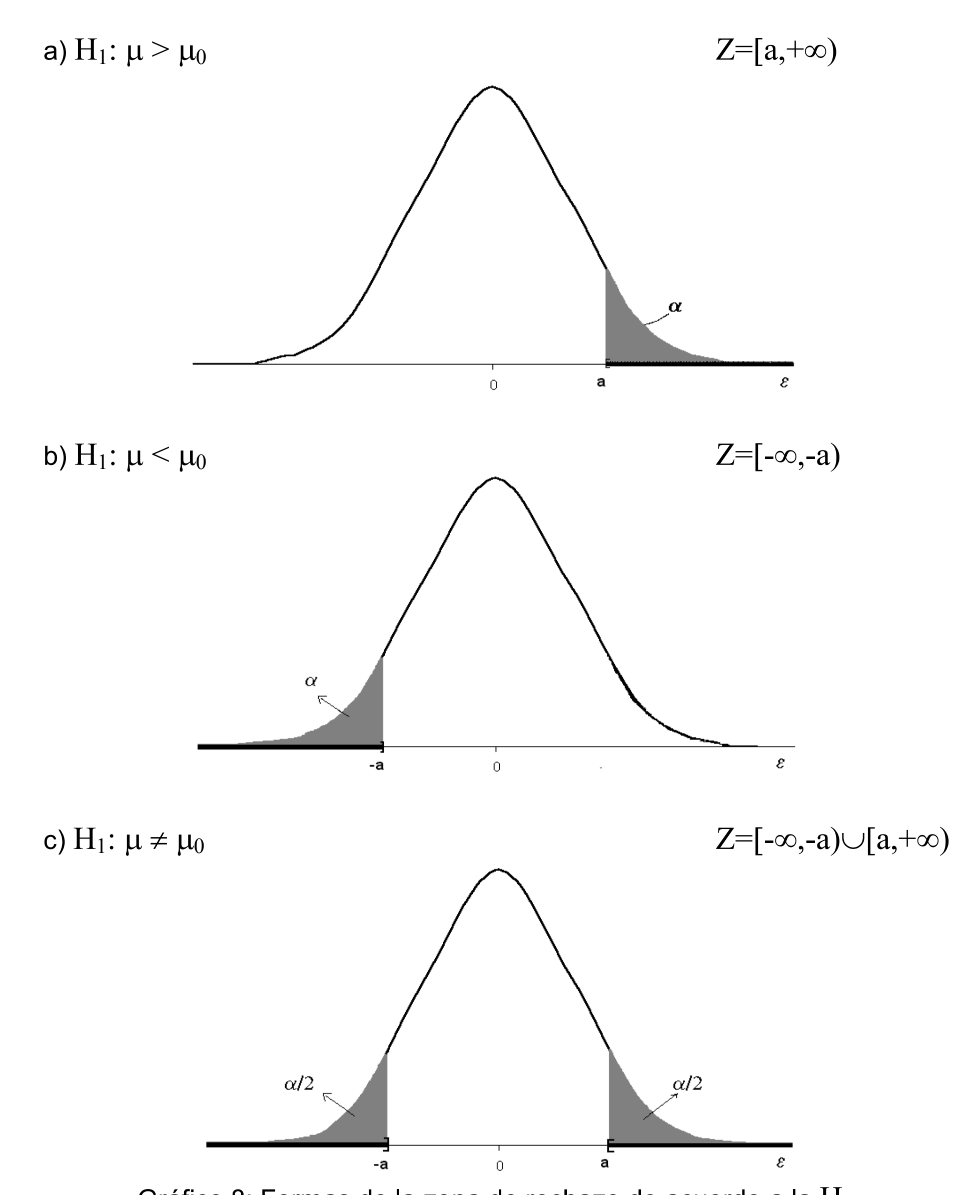


Gráfico 8: Formas de la zona de rechazo de acuerdo a la H_1

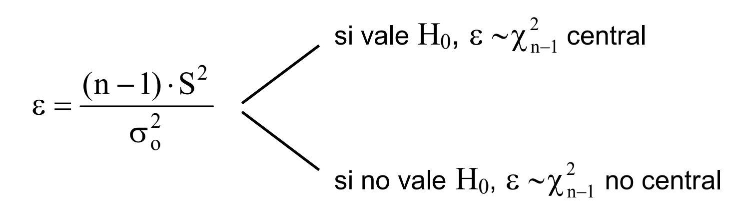
7.3 Prueba de hipótesis para la varianza de una Distribución Normal

Existen situaciones en la práctica donde el parámetro de interés es la varianza poblacional (por ejemplo si se desea conocer la variabilidad en instrumentos de medición). En otros casos, como el presentado al final de la Sección 7.2.2, para decidir qué estadístico usar en la prueba de hipótesis para la media se debe determinar si la varianza es distinta de un valor prefijado. Entonces, bajo el supuesto de la que la variable en estudio tiene distribución Normal, estos problemas se reducen a realizar una prueba para la varianza de una distribución Normal. Las hipótesis para este caso y el procedimiento a seguir son los siguientes.

Si el investigador sospecha que la varianza es un valor conocido σ_o^2, entonces las hipótesis a plantear son:

$$H_0: \sigma^2 = \sigma_o^2 \qquad H_1: \sigma^2 \neq \sigma_o^2$$

y el estadístico de contraste es



Como en las pruebas anteriores, prefijando el valor del nivel de significación α se determina la zona de rechazo Z, la cual en este caso particular es de la forma $Z=(0,a]\cup[b,+\infty)$ con a=$\chi^2_{n-1;1-\alpha/2}$ y b=$\chi^2_{n-1;\alpha/2}$. Gráficamente

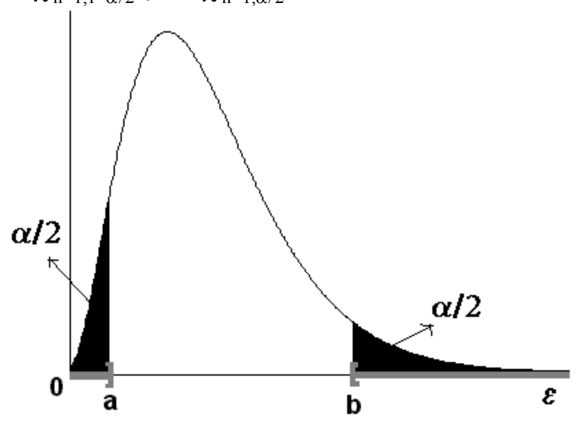


Gráfico 9: Zona de rechazo y nivel de significación

En base al valor numérico del estadístico y a la zona de rechazo determinada, la decisión y conclusión pueden ser:

- si se rechaza H_0 se puede afirmar que la varianza en cuestión es distinta al valor postulado σ_o^2 con probabilidad de cometer Error de Tipo I igual a α.
- si no se rechaza H_0 se puede decir que no hubo suficiente evidencia para afirmar que la varianza en cuestión es distinta al valor postulado σ_o^2, con probabilidad de cometer error de tipo II.

7.4 Prueba de hipótesis para la proporción de una Distribución Binomial para muestras de tamaños grandes

Retomando el Problema 6.1, suponga que se desea *determinar si la proporción* **p** *de animales afectados por aftosa en dicha zona es diferente de* 0.80. Para ello se seleccionan 120 animales y se cuenta el número de animales afectados por aftosa encontrándose 90 de ellos afectados.

Se sabe que la variable X: número de animales afectados en una selección, es una variable aleatoria que tiene distribución Bernoulli de parámetro **p** desconocido. Como en el Capítulo anterior, se define $Y = \sum_{i=1}^{120} X_i$ donde $Y \sim B(120,p)$.

Utilizando la relación entre las distribuciones Binomial y Normal, cuando n es grande, la prueba de hipótesis se puede resumir en los siguientes pasos:

1. H_0: $p=0.80$ $\qquad$ H_1: $p\neq 0.80$

2. $$\varepsilon = \frac{\hat{p} - p_o}{\sqrt{p_o(1-p_o)/n}}$$
Si vale H_0, $\varepsilon \sim N(0,1)$
Si no vale H_0, $\varepsilon \sim N(c,1)$

3. La zona de rechazo $Z=(-\infty,-a]\cup[a,+\infty)$. Si se fija un nivel de significación $\alpha=0.05$, se tiene que el valor critico es $a=1.96$.

4. Para determinar el valor numérico del estadístico se debe calcular

$\hat{p} = \frac{Y}{n} = \frac{90}{120} = 0.75$ y el $\text{Des}(\hat{p}) = \sqrt{0.80 \bullet \frac{0.20}{120}} = 0.0365$, luego el ε_c= -1.37

5. Como el ε_c no pertenece a la zona Z, entonces podemos concluir que no hay evidencias para decir que la proporción de animales afectados por aftosa es distinta de 0.80.

Para casos como éste, la hipótesis alternativa también puede ser unilateral (por mayor o menor) según lo que se desea probar.

7.5 Prueba de hipótesis para la diferencia de medias de dos distribuciones Normales

7.5.1 Muestras Independientes

En los ejemplos analizados hasta ahora se han estudiado situaciones en las cuales el interés era obtener información sobre las características numéricas de *una* distribución normal (media o varianza). En general ésta no será la situación que más comúnmente se presenta, ya que muchos problemas involucran más de una distribución normal, y por tanto se debe saber cómo comparar los parámetros en estos casos.

En cada una de las siguientes situaciones el investigador desea comparar las medias de *dos* distribuciones (cuando se tienen más de dos distribuciones en estudio se utiliza la técnica presentada en el capítulo siguiente).

1. Un gastroenterólogo desea comparar dos dietas para curar la úlcera péptica.

2. El delegado de salud pública compara niveles de inmunización de difteria en niños de dos secciones de una ciudad.

3. Un ingeniero agrónomo compara dos fertilizantes A y B aplicados sobre una variedad determinada de maíz.

4. Un veterinario desea comparar dos técnicas de congelación de embriones.

En cada una de las situaciones anteriores, midiendo *la variable* de interés se generan dos poblaciones estadísticas y la característica de interés en ambos casos es la $E(X)$ o media poblacional.

Para resolver este tipo de problemas se debe, como antes, suponer que la variable en estudio X tiene distribución normal. En las cuatro situaciones anteriores se desea comparar μ_1 versus μ_2.

En general, en estas pruebas H_0: μ_1 - μ_2 =0, mientras que H_1 puede plantearse de una (y sólo una) de las siguientes formas:

a) H_1: $\mu_1 - \mu_2 > 0$ b) H_1: $\mu_1 - \mu_2 < 0$ c) H_1: $\mu_1 - \mu_2 \neq 0$

Así por ejemplo en la Situación 3 el ingeniero agrónomo desea comparar el *rendimiento medio* de la variedad de maíz que fue sometido al fertilizante A versus el *rendimiento medio* de la variedad de maíz que fue sometida al fertilizante B.

Las hipótesis para probar lo deseado se plantean como:

$$H_0: \mu_A - \mu_B = 0 \qquad H_1: \mu_A - \mu_B \neq 0$$

las cuales indican

H_0: El rendimiento medio de la variedad de maíz tratada con el fertilizante A *es igual* al rendimiento medio de la variedad de maíz tratada con el fertilizante B .

H_1: El rendimiento medio de la variedad de maíz tratada con el fertilizante A *es diferente* al rendimiento medio de la variedad de maíz tratada con el fertilizante B .

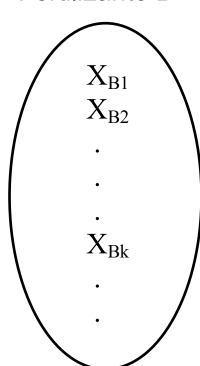
Para decidir si los dos fertilizantes producen el mismo efecto o no con respecto al rendimiento de la variedad de maíz en estudio se deben elegir n_A parcelas de maíz y tratarlas con el fertilizante A y n_B parcelas de maíz y tratarlas con el fertilizante B. Una suposición muy importante aquí es que las parcelas tratadas con uno y otro fertilizante son independientes.

Se puede representar la situación de la siguiente manera:

Población. Estadística Fertilizante A

X_{A1}
X_{A2}
.
.
.
X_{Ak}
.
.

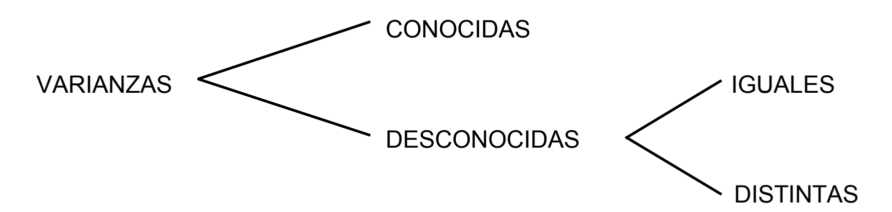
Población. Estadística Fertilizante B



Luego las características numéricas de estas dos poblaciones son

$E(X_A)=\mu_A$ y $Var(X_A)=\sigma_A^2$ $E(X_B)=\mu_B$ y $Var(X_B)=\sigma_B^2$

Llegado a este punto, se deben distinguir diferentes casos que surgen del conocimiento que se tenga acerca de las varianzas σ_A^2 y σ_B^2 que son:



En general, cuando se estudian dos distribuciones normales sucede que tanto la media como la varianza son desconocidas. Sin embargo para facilitar el desarrollo de la prueba se presentará el caso en que las varianzas son conocidas.

7.5.1.1 Varianzas poblacionales conocidas

Suponga que la variable rendimiento se distribuye como sigue $X_A \sim N(\mu_A, \sigma_A^2)$ y $X_B \sim N(\mu_B, \sigma_B^2)$ independientes donde se supone que σ_A^2 y σ_B^2 son conocidas pero μ_A y μ_B son desconocidas.

Para probar las hipótesis planteadas se define un estadístico que está basado en la diferencia de las medias muestrales generadas por cada una de las muestras aleatorias de tamaño n_A y n_B, o sea sobre la base de $(\overline{X}_A - \overline{X}_B)$. Dado que la media muestral es una variable aleatoria con distribución normal, se tiene para este caso que

$$\overline{X}_A \sim N\left(\mu_A, \sigma_A / \sqrt{n_A}\right) \qquad \text{y} \qquad \overline{X}_B \sim N\left(\mu_B, \sigma_B / \sqrt{n_B}\right)$$

Se puede probar que si $\overline{X}_A$ y $\overline{X}_B$ son dos variables aleatorias independientes, entonces la diferencia $(\overline{X}_A - \overline{X}_B)$ es una variable aleatoria con distribución normal de parámetros $(\mu_A - \mu_B)$ y $\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}$ (Mendenhall, W. et. al. - 1994).

Entonces el estadístico de contraste, cuando las varianzas son conocidas es

$$\varepsilon = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}}$$

Si vale H_0, $\varepsilon \sim N(0,1)$

Si no vale H_0, $\varepsilon \sim N(c,1)$

donde $c = \dfrac{(\mu_A - \mu_B)}{\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}}$.

7.5.1.2 Varianzas poblacionales desconocidas

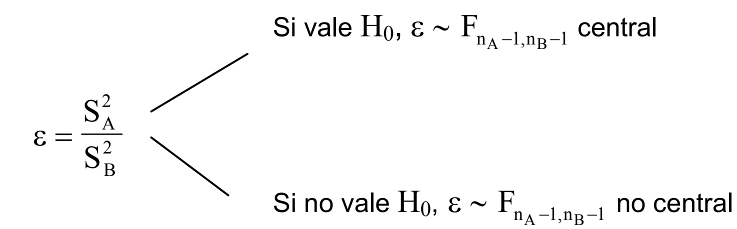
Cuando las varianzas poblacionales son desconocidas se presenta el problema de decidir si se las puede considerar

- *iguales*;
- *distintas*.

Para decidir esto se debe realizar la siguiente prueba *de homogeneidad de varianzas*.

$$H_o : \frac{\sigma_A^2}{\sigma_B^2} = 1 \qquad\qquad H_1 : \frac{\sigma_A^2}{\sigma_B^2} \neq 1$$

El estadístico de contraste es



Si vale H_0, $\varepsilon \sim F_{n_A-1,n_B-1}$ central

$$\varepsilon = \frac{S_A^2}{S_B^2}$$

Si no vale H_0, $\varepsilon \sim F_{n_A-1,n_B-1}$ no central

donde (n_A-1) y (n_B-1) indican los grados de libertad para el numerador y denominador de la distribución F de Fisher respectivamente.

La zona de rechazo para este caso toma la forma $Z=(0,a]\cup[b,+\infty)$ donde **a** y **b** se obtienen de tablas de la distribución F de Fisher. Gráficamente

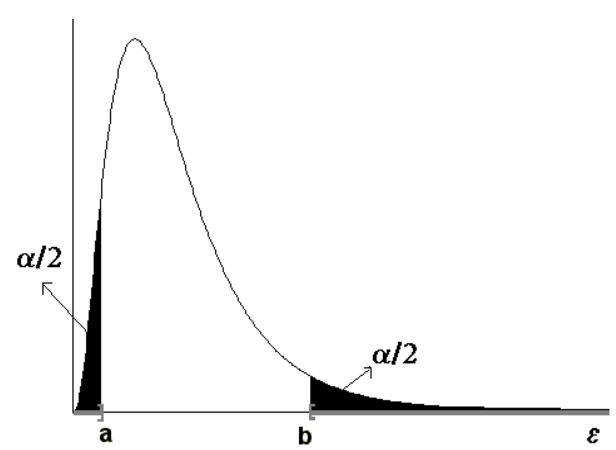


Gráfico 10: Zona de rechazo y nivel de significación

1. Si no se rechaza H_0 se puede asumir que las varianzas son iguales, entonces para la prueba de diferencia de medias el estadístico de contraste es:

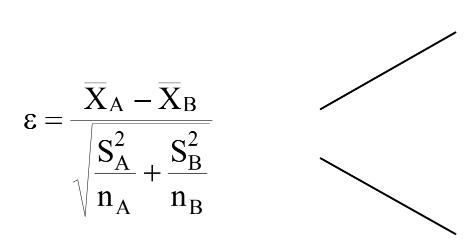
Si vale H_0, $\varepsilon \sim t_{n_A+n_B-2}$ central

$$\varepsilon = \frac{\overline{X}_A - \overline{X}_B}{S_D}$$

Si no vale H_0, $\varepsilon \sim t_{n_A+n_B-2}$ no central

donde $S_D = \sqrt{\dfrac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}}\sqrt{\dfrac{n_A + n_B}{n_A \bullet n_B}}$

2. Si se rechaza H_0 entonces el estadístico de contraste para la comparación de las medias de dos distribuciones es:



$$\varepsilon = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{\dfrac{S_A^2}{n_A} + \dfrac{S_B^2}{n_B}}}$$

Si vale H_0, $\varepsilon \sim t_f$ central

Si no vale H_0, $\varepsilon \sim t_f$ no central

donde $f = \dfrac{\dfrac{S_A^2}{n_A} + \dfrac{S_B^2}{n_B}}{\left(\dfrac{S_A^2}{n_A - 1}\right)^2 + \left(\dfrac{S_B^2}{n_B - 1}\right)^2} - 2$ se redondea al entero más cercano. Este es llamado el Método de Satterthwaite.

7.5.2 Muestras dependientes (apareadas)

En ciertos casos en los que se desea comparar medias de dos distribuciones es posible tomar muestras dependientes, que se pueden generar por dos situaciones

- La *misma* unidad experimental es medida en dos ocasiones diferentes.
- Las mediciones se obtienen a partir de unidades experimentales relacionadas (por ejemplo gemelos o cerdos de una misma camada, que tienen características genéticas comunes).

Ejemplos de las dos situaciones anteriores son:

1. Se mide el tono muscular de un grupo de individuos antes y después de un ejercicio severo.
2. Para comparar el efecto de dos antiparasitarios en cerdos se seleccionan pares de cerdos de la misma camada y se aplica uno de los medicamentos a uno de los cerdos del par y el otro al restante.

Generar muestras apareadas es lo ideal, pues de esta manera se eliminan las fuentes de variación que puedan existir entre unidades experimentales diferentes. Luego si existen diferencias en las medias poblacionales, éstas sólo podrán ser atribuidas al tratamiento en estudio.

Si los datos de la Muestra 1 se denotan X_{1i} y los de la Muestra 2 se denotan X_{2i} (i=1,2,...,n), las diferencias $d_i = X_{1i} - X_{2i}$ generan una nueva muestra que es la utilizada para la prueba.

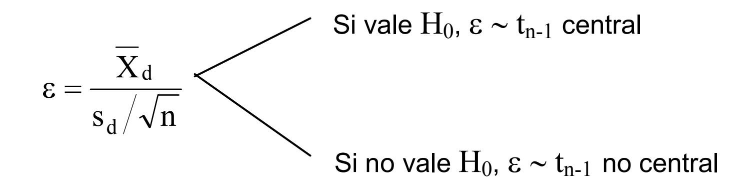
En estas pruebas la hipótesis nula es H_0: $\mu_d = 0$, mientras que la hipótesis alternativa H_1 puede plantearse de una (y sólo una) de las siguientes formas:

a) $H_1: \mu_d > 0$ b) $H_1: \mu_d < 0$ c) $H_1: \mu_d \neq 0$

donde $\mu_d = E(d_i) = E(X_{1i}-X_{2i}) = E(X_{1i})-E(X_{2i})$ o bien $\mu_d = \mu_1 - \mu_2$

Es decir, hacer inferencia acerca de la diferencia de las medias de dos tratamientos $(\mu_1 - \mu_2)$, es hacer inferencia acerca de la media de las diferencias (μ_d).

El estadístico de contraste es:



donde $\overline{X}_d$ y S_d son la media y el desvío estándar de la muestra de las diferencias d_i.

En la presente sección se presentaron sólo las hipótesis y el estadístico de contraste, dado que las pruebas de hipótesis siguen como en el caso de una media.

En las pruebas anteriores (para comparar dos medias o dos varianzas) se pueden plantear también hipótesis alternativas unilaterales, según lo que el investigador desee probar.

7.6 Prueba de hipótesis para la diferencia de proporciones de dos distribuciones Binomiales independientes

En muchas ocasiones, el interés recae en comparar las proporciones de ocurrencia de cierto suceso en dos grupos considerados por alguna razón diferentes.

Una situación de este tipo es la siguiente: “Se desea probar si la proporción de animales enfermos en dos regiones, consideradas geográficamente distintas, son estadísticamente diferentes”.

Para resolver este problema se tomaron al azar 400 animales de una de las zonas en estudio y se encontró que 190 de ellos estaban enfermos, en tanto que de la otra zona se tomaron al azar 800 animales de los cuales 300 estaban enfermos. Las proporciones muestrales para cada zona son:

$$\hat{p}_1 = \frac{190}{400} = 0.475 \quad \text{y} \quad \hat{p}_2 = \frac{300}{800} = 0.375 .$$

Ahora se van a plantear las hipótesis de interés

H_0: $p_1 - p_2 = 0$ $\qquad$ H_1: $p_1 - p_2 \neq 0$

En términos del problema:

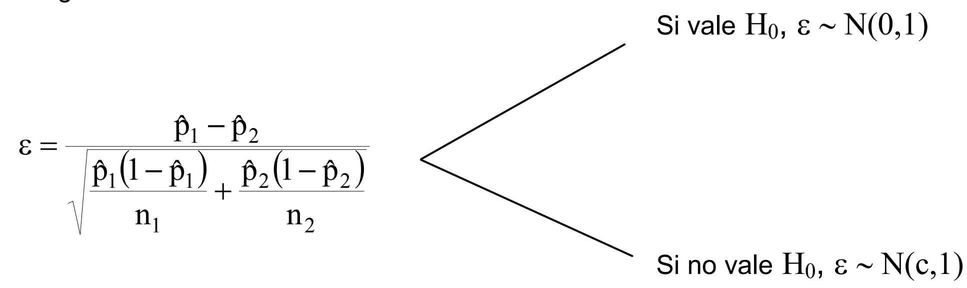
H_0: La proporción de animales enfermos en las dos zonas en estudio es la misma.

H_1: La proporción de animales enfermos en las dos zonas en estudio es diferente.

Si $X_1 \sim B(n_1,p)$, $X_2 \sim B(n_2,p)$ con n_1 y n_2 grandes (mayores a 30), por la relación entre las distribuciones Normal y Binomial, la variable aleatoria $\hat{p}_1 - \hat{p}_2$ tiene distribución

normal con $E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$ y $Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$. (Para verificar estas afirmaciones utilizar las propiedades de Esperanza y Varianza.)

Luego el estadístico de contraste resulta



$$\varepsilon = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

Con $\hat{p}_1$ y $\hat{p}_2$ estimadores de p_1 y p_2.

Si se toma un nivel de significación del 0.05 el valor crítico es $a=1.96$, luego la zona de rechazo que resulta es $Z=(-\infty,-1.96]\cup[1.96,+\infty)$. En este caso el $\varepsilon_c=3.31$, por lo que se rechaza la hipótesis nula, con probabilidad de cometer error de tipo I.

Conclusión: *Las proporciones de animales enfermos en las dos zonas en estudio son diferentes*.

La hipótesis alternativa para otras situaciones también podría plantearse como

H_1: $p_1 - p_2 > 0$ ó H_1: $p_1 - p_2 < 0$

Para cada una de las pruebas presentadas en este capítulo se pueden construir intervalos de confianza

7.7 Relación entre Intervalo de Confianza y Prueba de Hipótesis.

Ahora será presentada una forma alternativa de tomar una decisión en una prueba de hipótesis bilateral.

Para ello suponga que se plantean las siguientes hipótesis acerca de la media de una población normal.

H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$

Para tomar una decisión acerca de éstas hipótesis se pueden seguir dos caminos, tal como se indica a continuación:

a) Realizar la prueba de hipótesis construyendo, para un valor de α especificado, la zona de rechazo $Z=(-\infty,-a]\cup[a,+\infty)$, de donde la zona de no rechazo es $(-a,a)$.

b) Construir un intervalo de confianza para μ de nivel $(1-\alpha)$, el cual según lo presentado en el Capítulo 6 tiene la forma $\left(\overline{X} - \frac{a\sigma}{\sqrt{n}}, \overline{X} + \frac{a\sigma}{\sqrt{n}}\right)$.

Si μ_0 pertenece al intervalo de confianza, significa que $\overline{X} - \frac{a\sigma}{\sqrt{n}} < \mu_o < \overline{X} + \frac{a\sigma}{\sqrt{n}}$. A partir de esta expresión y siguiendo los pasos algebraicos inversos a los realizados para construir el intervalo de confianza se obtiene que

$$-a < \frac{\overline{X} - \mu_o}{\sigma / \sqrt{n}} < a \text{ , o equivalentemente } -a < \varepsilon < a$$

(donde ε es el estadístico de contraste de la prueba). La última desigualdad indica que el estadístico pertenece a la zona de no rechazo de la prueba, por lo que no se rechaza H_0.

Esto dice que a un valor de μ_0 que pertenece al intervalo de confianza de nivel $(1-\alpha)$ le corresponde un valor del estadístico que lleva a no rechazar la hipótesis nula en la prueba bilateral de nivel α. Por otro lado si μ_0 no pertenece al intervalo de confianza, el estadístico de contraste caerá en la zona de rechazo.

En conclusión para poder utilizar la relación entre una prueba de hipótesis y un intervalo de confianza se deben verificar las siguientes condiciones:

- La prueba debe ser bilateral, esto es H_1: $\mu \neq \mu_0$
- El nivel de confianza del intervalo debe ser el complemento del nivel de significación de la prueba.

Resultados análogos son válidos para todos las pruebas presentadas y los intervalos de confianza correspondientes.

7.8 Comentarios finales

Para finalizar es conveniente realizar una reflexión a cerca de las suposiciones sobre las cuales se han construido los intervalos de confianza y las diferentes pruebas de hipótesis. En este punto es claro que los procedimientos estadísticos de inferencia proporcionan un camino objetivo y veraz para conocer características poblacionales en base a la información dada por muestras aleatorias. Estos procedimientos en general son válidos si la variable en estudio tiene distribución Normal (o aproximadamente Normal). Los enfoques fortuitos y casuales para la aplicación de los métodos estadísticos, sin una comprensión de sus suposiciones y de las posibles consecuencias si éstas no se satisfacen, muchas veces llevan a una mala interpretación y a conclusiones erróneas.

Como ya se ha visto la distribución t de Student juega un papel muy importante para formular inferencias con respecto a las medias, en forma especial en muestras de tamaño pequeño. La distribución t de Student se basa en la suposición de que el muestreo se realiza sobre una distribución normal, pero si esto no fuera así, el uso de esta distribución es incorrecto (la región crítica determinada para un α dado, resulta de probabilidad diferente que el valor especificado).

Otro punto importante es que la inferencia se basa en el hecho que se utilizan muestras aleatorias, es decir observaciones provenientes de variables aleatorias independientes idénticamente distribuidas. Si esta suposición no se verifica, es probable que, cualquier inferencia estadística que se realice sea errónea sin importar el tamaño de la muestra.

Ejercicios de Aplicación

1.

Se sospecha que una máquina embotelladora de leche no funciona adecuadamente. El volumen promedio de leche de las botellas debe ser de 970 cm^3. Se supone que el desvío estándar de la variable "volumen" asume un valor de 20 cm^3.

a) ¿Cuál es el objetivo del estudio?. En el marco de la teoría de las pruebas de hipótesis, ¿cómo plantearía este objetivo?.

b) Para las hipótesis planteadas en el inciso anterior, ¿qué estadístico de contraste se debe usar?.

c) Hallar la zona de rechazo para un nivel de significación de 0.05.

d) Para poder tomar una decisión respecto a las hipótesis planteadas se tomaron al azar 9 botellas, encontrándose una media de 977 cm^3. ¿Hay evidencia estadística para concluir que la máquina funciona mal?.

e) ¿Cuál sería la conclusión si se hubiesen tomado 49 botellas obteniendo también un volumen medio de 977 cm^3?

2.

Se sabe que ciertas ratas con una alimentación habitual tiene una ganancia de peso medio de 65 gr. durante los tres primeros meses de vida. Para probar el efecto de una nueva dieta, se alimentaron 30 ratas desde el nacimiento hasta la edad de tres meses, encontrándose un aumento medio de peso de 70.75 gr. y una varianza de 10 $gr.^2$. ¿ Hay evidencias estadísticamente significativas para sostener, al nivel del 1%, que la nueva dieta aumenta la ganancia de peso promedio?.

3.

Un establecimiento dedicado a la elaboración de alimentos balanceados afirma que su producto en aves de una cierta raza y de un mes de vida produce un aumento medio de peso mayor a 100 gr. por semana.

a) ¿Cuáles son las hipótesis a contrastar?. Estadísticamente e interpretarlas en términos del problema.

b) Para tomar una decisión sobre las hipótesis planteadas se eligieron al azar 16 aves de esa raza y se les suministró el alimento balanceado durante una semana, obteniéndose con estos datos un valor de ε_C=2.13. Realizar el análisis correspondiente y establecer conclusiones usando el valor p.

4.

Para cada una de las siguientes situaciones plantear las hipótesis a contrastar y, de acuerdo al valor p obtenido en base a una muestra de la población en estudio, responder a las preguntas que se formulan:

a) Investigaciones anteriores han determinado que la duración media de sobrevida de los pacientes afectados por cierta enfermedad es de 3.4 meses. Un investigador afirma que una nueva droga prolonga la vida de estos pacientes. Para p=0.006, ¿qué se puede afirmar sobre el efecto de la nueva droga?.

b) Un inspector del Instituto Nacional de Tecnología Agropecuaria sospecha que el contenido medio de semillas de un cierto producto agrícola es inferior al indicado en la

etiqueta (que es de 45 gr.). Si el valor p=0.075, ¿sancionará el inspector al establecimiento que fabrica y envasa dicho producto?.

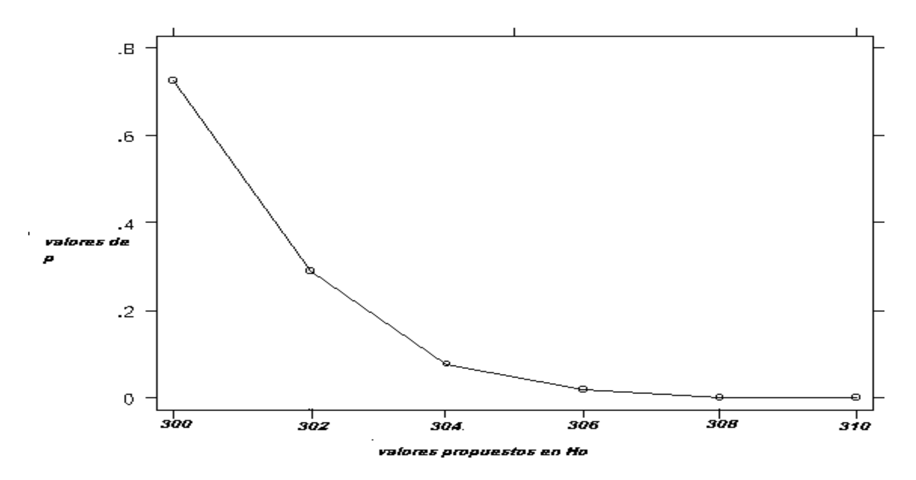
c) En una investigación se midió el tiempo de reacción en segundos a un estímulo a un grupo de animales de cierta raza. Se afirma que el tiempo medio de reacción es inferior a los 50 segundos. Para un valor p=0.32, ¿a qué conclusión se llega respecto del tiempo de reacción promedio?

5.

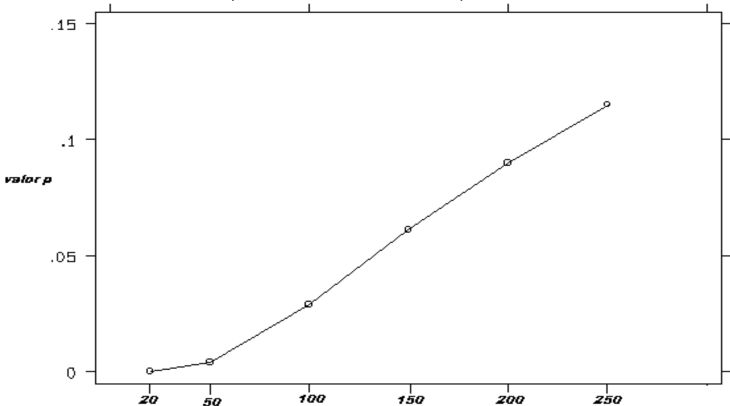
Un veterinario desea estudiar el efecto de un antiparasitario en cerdas después de la parición. Para ello selecciona 25 cerdas afectadas, las trata con dicho medicamento y les mide el tiempo de recuperación en días encontrándose un tiempo de recuperación promedio de 2.33 días y un desvío estándar de 1.2 días. Dicho profesional sostiene que si el tiempo medio de recuperación es inferior a los 3 días es posible considerar que el medicamento produce el efecto esperado.
¿Produce el antiparasitario el efecto esperado con una probabilidad máxima de error de tipo uno de 0.1 ?.

6.

a) Se quieren contrastar las hipótesis H_o: $\mu=\mu_0$ versus H_1: $\mu\neq\mu_0$, donde μ representa la media de una variable, X, con distribución normal. Se obtiene una muestra de n valores de X, con una media de 299. En la figura que se presenta a continuación se muestran los valores de p para diferentes valores de μ_0, manteniendo siempre fijo $\overline{X}$. Discutir el comportamiento del valor p como función de μ_0.



b) Se extraen varias muestras de una población, suponga que la media muestral es 299 en cada una de ellas. En la figura de abajo se muestran los diferentes valores de **p** para contrastar las hipótesis H_o: $\mu=305$ versus H_1: $\mu\neq305$:



¿Cuál es el efecto de la varianza muestral sobre el valor **p**?.

7.

Lotes de 9 abejas fueron alimentados en comederos con jarabes de, concentración C_1=20% y C_2 =65% a una milla del panal. Al llegar al panal se les separa el saco de miel determinándose la concentración de líquido. En cada caso se midió la disminución de dicha concentración en relación con la del comedero, registrándose para la concentración C_1 una media de 0.49% y para la C_2 una media de 1.9%.
¿Se puede sostener que fue mayor la disminución en la concentración C_2 , suponiendo que los desvíos estándares poblacionales son 1.09 % y 1.3 % para cada concentración respectivamente, para una máxima probabilidad de error de tipo I del 0.10 ?

8.

En un establecimiento avícola se desea comparar los efectos de dos dietas sobre el peso de pollitos de la misma raza y edad. Suponer que la variable peso tiene distribución Normal.

a) ¿Cuál es la variable en estudio y a qué tipo corresponde?

b) ¿Cuántas poblaciones estadísticas se generan y qué distribución tiene la variable en cada una de ellas?

c) ¿Cuáles son las hipótesis a contrastar?.

d) Para tomar una decisión sobre las hipótesis planteadas se tomaron 18 pollitos al azar, suministrándosele cada dieta a 9 pollitos y luego de cierto lapso de tiempo se determinó su peso (en gr.), arrojando los siguientes valores:

Dieta									
A	10	10	13	12	16	9	15	12	13
B	15	16	11	16	17	10	14	12	15

Usar los valores de p que se muestran en la siguiente tabla para responder

Prueba	Valor p
Homogeneidad de Varianzas	0.8941
Igualdad de Medias (suponiendo varianzas iguales)	0.1345
Igualdad de Medias (suponiendo varianzas distintas)	0.1351

e) ¿Hay evidencia para sostener que las varianzas para ambas poblaciones son distintas?

f) ¿Existe diferencia entre los pesos medios para ambas poblaciones ?

9.

Con el objetivo de comparar el contenido total de nitrógeno (en gramos por 100 cm^3) en plasma de ratas albinas normales en distintos momentos de vida, se seleccionaron 6 ratas de 37 días y 6 de 180 días y se determinó el contenido de nitrógeno obteniéndose los siguientes valores:

Tratamiento						
37 días	0.98	0.83	0.99	0.86	0.90	0.91
180 días	1.20	1.18	1.33	1.21	1.2	1.07

Utilizando los resultados de la tabla dada a continuación, ¿hay evidencia estadísticamente significativa para afirmar que los contenidos medios totales de nitrógeno a los 37 días y a los 180 días son diferentes?.

Prueba	Valor p
Homogeneidad de Varianzas	0.5786
Igualdad de Medias (suponiendo varianzas iguales)	0.0001
Igualdad de Medias (suponiendo varianzas distintas)	0.0001

10.

En un experimento se desea determinar si el contenido de hemoglobina en la sangre de perros cambia al aplicar un tratamiento con niacina. Para ello se tomaron 8 perros, y se midió el contenido de hemoglobina antes y después del tratamiento, obteniéndose las siguientes mediciones:

Trat. / Perro	1	2	3	4	5	6	7	8
Antes	12.6	12.6	13.7	11.1	11.3	12.2	10	11
Después	10.4	11.5	13.6	12.0	19.3	8.8	9.4	10.7

¿Modifica el tratamiento el contenido medio de hemoglobina?

11.

Dos raciones alimenticias van a compararse con respecto a su efecto en el incremento de peso en cerdos. Para ello se seleccionaron pares de cerdos de la misma camada, suministrándole a cada animal del par una ración diferente. Los incrementos de peso se muestran a continuación.

Ración					
1	0.454	0.908	1.816	2.27	3.362
2	0.816	1.362	4.086	4.54	4.086

¿ Se puede afirmar (en sentido estadístico) que la ración 2 es más efectiva que la ración 1?

8 Introducción al Análisis de la Varianza

Objetivo:

- Conocer la técnica que permite comparar las medias de dos o más poblaciones y sobre la cual se basa el Diseño de Experimentos.
- Comprender la utilidad de la técnica cuando se desean comparar más de dos medias.

8.1 Introducción

En el Capítulo 7 se presentó la prueba para comparar las medias de dos poblaciones normales (diferencias de medias). En muchas situaciones es necesario comparar más de dos medias. En estos casos no es conveniente efectuar todas las comparaciones posibles de las medias tomadas de dos en dos, ya que al realizarlo de esta manera el nivel de significación verdadero no es el prefijado para la prueba, lo que puede llevar a conclusiones erróneas. Lo adecuado es estudiar *simultáneamente* las diferencias entre las medias de todas las poblaciones. Problemas como éstos se pueden resolver usando una importante técnica estadística conocida como *Análisis de la Varianza* (ANOVA)[1], método que fue desarrollado por R.A. Fisher.

Algunas situaciones en las que se desea comparar más de dos medias son:

- Un veterinario desea comparar el efecto de tres dietas de engorde en pollos parrilleros.

- Un ingeniero agrónomo desea comparar el efecto de dos fertilizantes y un control, para una cierta variedad de trigo.

- Un médico desea comparar el efecto de cuatro drogas para el dolor de cabeza.

Si bien el Análisis de la Varianza se puede considerar como una generalización de la prueba de diferencias de medias, el mecanismo en sí es muy distinto, pues la técnica de ANOVA está basada en la comparación de varianzas y no de medias.

Al estudiar la variación de un conjunto de datos cualesquiera, utilizando como medida de dispersión la varianza, no se han tenido en cuenta las causas de dicha variación. En muchos casos la variabilidad total existente es el resultado de más de una causa. La técnica de *Análisis de la Varianza* consiste en descomponer la variabilidad total en una variación atribuible a causas conocidas (efectos producidos por distintas dietas, fertilizantes o drogas) comúnmente denominadas *tratamientos* y en otra debida a causas desconocidas que no pueden ser controladas por el experimentador y son atribuibles al azar. Esta última es considerada como la variación intrínseca a la unidad experimental.

[1] ANOVA: sigla en inglés para Analysis of Variance

8.2 Análisis de la Varianza

A continuación se presenta una aplicación de esta técnica.

Ejemplo 1: Se desea determinar el efecto del estrés en ratas albinas. Para ello se midieron diferentes variables, entre ellas la ingesta de agua (en ml.) bajo tres tratamientos diferentes: Comida ad-libitum (Tratamiento 1), Comida restringida (Tratamiento 2) y Comida ad-libitum con estrés (Tratamiento 3). El objetivo de esta experiencia es determinar si la ingesta promedio de agua es diferente en ratas albinas sometidas a estos tres tratamientos

8.2.1 Modelo lineal

Cuando se está en una situación experimental como la del Ejemplo 1, la respuesta puede ser descripta con el *modelo de posición*

$$X_{ij} = \mu_i + \varepsilon_{ij} \qquad i=1,2,...,t\ ;\ j=1,2,...,n_i$$

donde :

X_{ij}: ingesta de agua de la **j**-ésima rata sometida al **i**-ésimo tratamiento;

μ_i: ingesta media de agua con el tratamiento **i** -ésimo;

ε_{ij}: componente del error aleatorio. Los ε_{ij} son variables aleatorias independientes distribuidas normalmente, con media cero y varianza común σ^2;

t: número de tratamientos;

n_i : cantidad de unidades experimentales asignadas al tratamiento **i**.

Equivalentemente se puede utilizar el *modelo de efectos*

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad i=1,2,...,t\ ,\ j=1,2,...,n_i$$

donde

μ: la ingesta media de agua.

τ_i: efecto no aleatorio del **i**-ésimo tratamiento. Con la restricción $\sum_{i=1}^{t} \tau_i = 0.$

Estos dos modelos son equivalentes porque $\mu_i = \mu + \tau_i$.

Las hipótesis pueden escribirse usando el modelo de posición como

$H_0: \mu_1 = \mu_2 = ... = \mu_t$ $\qquad$ H_1: al menos un μ_i diferente, i=1,2,...,t

y usando el modelo de efectos como

$H_0: \tau_1 = \tau_2 = ... = \tau_t = 0$ $\qquad$ H_1: al menos un τ_i diferente, i=1,2,...,t

8.2.2 Prueba de hipótesis

Como el objetivo de esta experiencia es determinar si la ingesta promedio de agua es diferente en ratas albinas sometidas a estos tres tratamientos. Esto puede ser planteado de la

siguiente manera:

$H_0: \mu_1 = \mu_2 = \mu_3$ $\qquad$ H_1: al menos un μ_i diferente, i=1,2,3

que se puede expresar:

H_0: La ingesta promedio de agua es la misma bajo los tres tratamientos.
H_1: La ingesta promedio de agua es diferente para al menos uno de los tratamientos.

Algunos elementos importantes para realizar al análisis estadístico son:

Experimento Aleatorio: Seleccionar al azar 24 ratas, dividirlas (aleatoriamente) en tres grupos de 8 ratas cada uno, para luego asignar un tratamiento aleatoriamente a cada una de los grupos.

Unidad Experimental: Una rata albina de peso inicial 220 gr. de un determinado sexo y edad.

Población de Unidades: Todas las ratas albinas de peso inicial 220 gr. de un determinado sexo y edad.

Muestras de unidades: 24 ratas albinas seleccionadas aleatoriamente para el estudio.

Variable en estudio: Ingesta de agua

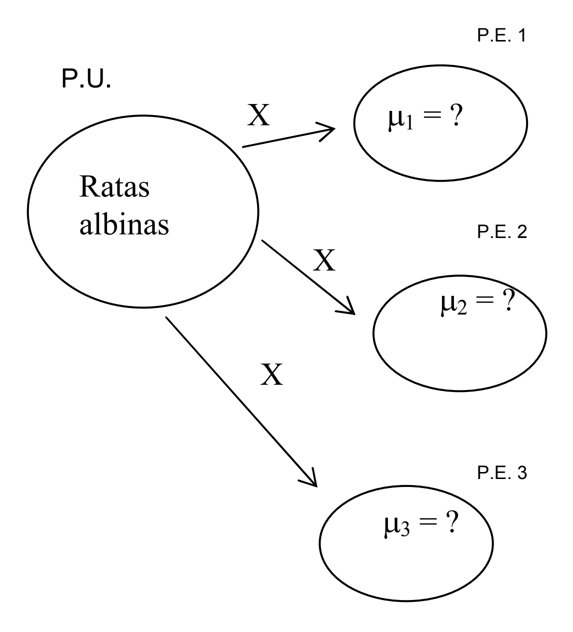
Población estadística: Se generan tres poblaciones estadísticas:

Población estadística 1: Todos los valores de ingesta de agua correspondientes a cada una de las ratas que reciben el Tratamiento 1.

Población estadística 2: Todos los valores de ingesta de agua correspondientes a cada una de las ratas que reciben el Tratamiento 2.

Población estadística 3: Todos los valores de ingesta de agua correspondientes a cada una de las ratas que reciben el Tratamiento 3.

Gráficamente:



Muestra Estadística: Análogamente se tienen tres muestras estadísticas, cuyos valores se indican en la Tabla 1.

Tabla 1: Ingesta de agua (en ml.) de ratas albinas sometidas a tres tratamientos

	Tratamiento 1	Tratamiento 2	Tratamiento 3
	3.80	3.10	3.84
	0.61	4.39	4.05
	0.81	3.04	1.75
	0.81	0.25	2.95
	1.49	0.38	5.62
	7.56	5.20	5.99
	3.00	3.60	4.60
	0.70	2.40	3.20
$\sum_{j=1}^{8} X_{ij}$	18.78	22.45	32.00
X_{ij}	2.35	2.81	4.00

Total General $X = \sum_{i=1}^{3}\sum_{j=1}^{8} X_{ij} = 73.23$ N = 24

Es importante destacar que este tipo de análisis permite el estudio de *una* sola variable, medida en situaciones (tratamientos) diferentes, las cuales generan sendas poblaciones estadísticas (una por cada tratamiento).

La técnica de Análisis de la Varianza consiste en descomponer a la variación total en una variación debida al efecto de los tratamientos (variación entre tratamientos) y una variación atribuida al azar (variación dentro de tratamiento). A continuación se muestra como obtener esta descomposición.

La variabilidad de un conjunto de datos respecto a su media se mide mediante la varianza, la cual se puede pensar como el cociente entre *la suma de cuadrados de los desvíos*, denominada en este contexto Suma de Cuadrados Total (SC_T) y la cantidad de datos menos 1. Simbólicamente se tiene

$$S^2 = \frac{\sum_{i=1}^{3}\sum_{j=1}^{8}\left(X_{ij} - \overline{\overline{X}}\right)^2}{24-1} = \frac{SC_T}{24-1}$$, donde $\overline{\overline{X}} = \frac{\sum_{i=1}^{3}\sum_{j=1}^{8} X_{ij}}{24}$ es la media general, X_{ij} es la j-ésima

observación en la i-ésima muestra y 24 es el número total de observaciones.

Es importante tener en cuenta en este punto que S^2 es un estimador de la varianza poblacional σ^2, y sobre la base de este hecho se construyen estimadores para la *variación total*, la *variación entre tratamientos* (muestras) y la *variación dentro de tratamiento*.

El valor de la *variación total* indica cuánto se apartan las observaciones X_{ij} de la media general $\overline{\overline{X}}$. Para esto, se considera al total de los datos (en este caso 24) como una sola muestra, entonces esta variación incluye tanto la producida por los tratamientos como la debida al azar.

Para determinar la *variación entre tratamientos se* toma un representante de cada muestra $\overline{X}_i$ y se considera como si hubiera una muestra de tres datos: $\overline{X}_1, \overline{X}_2, \overline{X}_3$. Luego, esta variación mide cuánto se apartan, en promedio, las medias de cada tratamiento de la media general de los 24 datos.

El valor de la *variación dentro de tratamiento* mide la variación dentro de cada muestra,

es decir, cuánto se aleja cada dato X_{ij} de $\overline{X}_i$. Esta variación recibe el nombre de *error experimental* .

A continuación se presentan las expresiones de estas tres fuentes de variabilidad, para el caso general en el que se comparan **t** tratamientos, cada uno con $\mathbf{n_i}$ unidades experimentales (con i=1, 2,...,t).

La Variación Total es

$$\frac{SC_T}{N-1} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{\overline{X}}\right)^2}{N-1}$$

la suma de cuadrados Total se puede dividir en dos sumandos, es decir

$$SC_T = \sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{\overline{X}}\right)^2 = \sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2 + \sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(\overline{X}_i-\overline{\overline{X}}\right)^2$$

expresión que se obtiene operando algebraicamente luego de sumar y restar $\overline{X}_i$. Los sumandos de la expresión anterior son la Suma de cuadrados dentro de Tratamiento (SC_{ee}) y la Suma de cuadrados entre Tratamientos (SC_t), respectivamente. Entonces:

$$SC_{ee} = \sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2 \qquad SC_t = \sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(\overline{X}_i-\overline{\overline{X}}\right)^2$$

y

$$SC_T = SC_{ee} + SC_t$$

Para comparar la variabilidad entre tratamientos con la variabilidad dentro de tratamiento se divide cada una de estas sumas por sus respectivos grados de libertad, obteniendo los llamados Cuadrados Medios

Cuadrado Medio de Error = $CM_{ee} = \dfrac{SC_{ee}}{N-t} = \sum_{i=1}^{t}\dfrac{\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2}{n_i-1}$

Cuadrado Medio entre Tratamientos = $CM_t = \dfrac{SC_t}{t-1} = \dfrac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(\overline{X}_i-\overline{\overline{X}}\right)^2}{t-1}$

Los cuadrados medios son dos estimadores de la varianza poblacional σ^2. Notar que ambos son variables aleatorias independientes.

Si las medias de las poblaciones en realidad son diferentes, se espera que las medias de las muestras difieran *mucho* una de otra. En ese caso habría mucha dispersión entre ellas y por lo tanto el CM_t sería un número grande con respecto al CM_{ee}. Por otro lado si las medias poblacionales son iguales (no hay efecto de tratamiento), se espera que las medias de las muestras sean muy parecidas y la variación existente entre ellas sólo se deba al azar; por lo tanto el CM_t estaría muy próximo al CM_{ee}. De acuerdo a este razonamiento parece natural que el cociente entre estos cuadrados medios permita determinar si hay efecto de tratamiento, o equivalentemente si las medias poblacionales son distintas.

Por lo tanto el estadístico adecuado para esta prueba está dado por el cociente entre

los cuadrados medios y tiene una distribución F-Fisher [2]

Si vale H_0, $F \sim F_{t-1,N-t}$ central

$$\varepsilon = \frac{CM_t}{CM_{ee}}$$

Si no vale H_0, $F \sim F_{t-1,N-t}$ no central

En realidad el cuadrado medio entre tratamientos (CM_t) estima a la varianza poblacional más el efecto de tratamientos ($\sigma^2+\tau_i$) y el cuadrado medio dentro de tratamiento (CM_{ee}) es un estimador de la varianza poblacional (σ^2) lo que permite reescribir el estadístico de contraste de la siguiente manera

$$\varepsilon = \widehat{\frac{\sigma^2 + n\frac{\sum_{i=1}^{t}\tau_i}{t}}{\widehat{\sigma^2}}}$$

Si este cociente es:

- cercano a 1 no hay efecto de tratamiento, lo que llevaría a *no rechazar* H_0.
- mayor que 1 hay efecto de tratamiento, lo que llevaría a *rechazar* H_0.
- menor a 1, se debería *no rechazar* H_0. En este caso, como el CM_{ee} es un número grande habría que repetir la experiencia pues esto podría indicar la existencia de variaciones no controladas.

Hasta este punto se han planteado las hipótesis y el estadístico. A continuación se determina la zona de rechazo que, por las consideraciones realizadas acerca de los valores que puede tomar el estadístico, tiene la forma $Z=[a,+\infty)$. En general, la zona de rechazo y el nivel de significación tienen la siguiente forma:

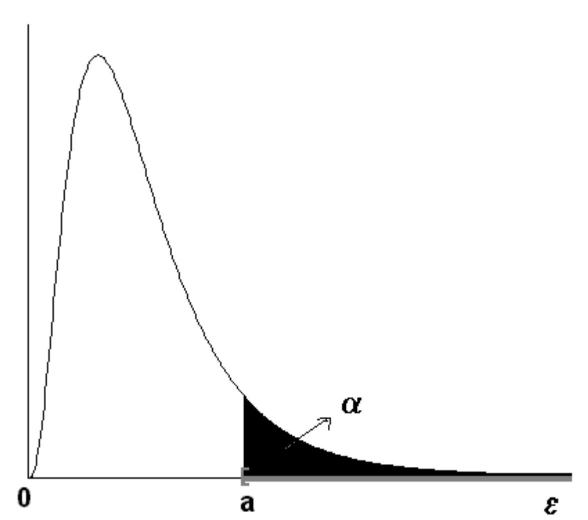


Gráfico 1: Nivel de significación y zona de rechazo.

Para el Ejemplo considerado t=3, N=24, n_i=8 . Si se decide prefijar un nivel de

[2] Esto se debe a que los cuadrados medios son variables aleatorias independientes con distribución χ^2 y el cociente de dos variables con distribuciones χ^2 dividido sus grados de libertad tiene distribución F de Fisher (Capítulo 5).

significación α=0.05, el valor crítico obtenido de la Tabla F del Apéndice es $F_{2,21,0.05}=3.47$ entonces $Z=[3.47,+\infty)$.

Antes de determinar el valor numérico del estadístico de contraste para el problema planteado se muestran las fórmulas de trabajo de las sumas de cuadrados total, entre tratamientos y dentro de tratamientos.

Fórmulas de Trabajo

$$SC_T = \sum_{i=1}^{t}\sum_{j=1}^{n_i} X_{ij}^2 - \frac{X^2}{N} \text{ , donde } X = \sum_{i=1}^{t}\sum_{j=1}^{n_i} X_{ij}$$

$$SC_t = \sum_{i=1}^{t}\frac{X_i^2}{n_i} - \frac{X^2}{N} \text{ , donde } X_i = \sum_{j=1}^{n_i} X_{ij}$$

$$SC_{ee} = SC_T - SC_t$$

Para el problema planteado

$$SC_T = (3.8)^2 + (0.61)^2 + \cdots + (3.2)^2 - \frac{(73.23)^2}{24} = 88.35$$

$$SC_t = \frac{(18.78)^2}{8} + \frac{(22.45)^2}{8} + \frac{(32)^2}{8} - \frac{(73.23)^2}{24} = 11.64$$

$$SC_{ee} = 88.36 - 11.64 = 76.71$$

Esta información se resume en una tabla llamada “Tabla de Análisis de la Varianza” (Tabla ANOVA).

Tabla 2: Tabla ANOVA para los datos del Ejemplo 1

Fuentes de Variación	*g.l.*	*S.C.*	*C.M.*	*F*
Tratamiento	2	11.64	5.82	1.593
Error Experimental	21	76.71	3.65	
Total	23	88.35		

Como $\varepsilon \notin Z$ no hay evidencias para rechazar H_0 con probabilidad de error de tipo II.

Conclusión: *no hay evidencias suficientes para decir que la ingesta de agua promedio de todas las ratas fue diferente para alguna de las tres comidas* (tratamientos).

Es usual expresar esta conclusión como *no se detectan diferencias significativas entre las tres comidas* (que significa no haber rechazado H_0).

8.2.3 Pruebas a Posteriori

Si la decisión hubiese sido rechazar la hipótesis nula, se podría concluir que *la ingesta de agua promedio de todas las ratas fue diferente para alguna de las tres comidas* o

equivalentemente se expresa como *hay diferencias significativas entre las tres comidas* (que significa haber rechazado H_0).

En esta última situación interesa determinar cuál o cuáles son los tratamientos que producen la diferencia. Para ello se puede recurrir a otras pruebas llamadas *Pruebas a Posteriori*. Éstas permiten realizar las comparaciones entre los tratamientos que intervienen en la experiencia para detectar entre quienes está la diferencia. Reciben la denominación de "a posteriori" porque se realizan después de rechazar la hipótesis de que las medias de todos los tratamientos son iguales.

Las más usadas son las pruebas de Tukey, Dunnet, Scheffé, etc. Ellas pueden ser consultadas en Steel, R.G.D. y Torrie, J.H. (1985).

Para el problema planteado, en caso de que la conclusión hubiese sido que alguna de *las ingestas medias de agua es diferente*, las hipótesis correspondientes a la prueba de Tukey son:

H_0: $\mu_1 = \mu_2$ $\qquad$ H_1: $\mu_1 \neq \mu_2$

H_0: $\mu_1 = \mu_3$ $\qquad$ H_1: $\mu_1 \neq \mu_3$

H_0: $\mu_2 = \mu_3$ $\qquad$ H_1: $\mu_2 \neq \mu_3$

Otra forma de detectar cuáles son los tratamientos que producen efectos diferentes es usar intervalos de confianza para las diferencias de medias $(\mu_1 - \mu_2)$, $(\mu_1 - \mu_3)$, $(\mu_2 - \mu_3)$.

8.2.4 Supuestos para la validez del modelo

Para que las conclusiones obtenidas del análisis sean válidas se requiere que se cumplan ciertos supuestos, que se pueden resumir de la siguiente manera:

1. Las muestras deben obtenerse aleatoriamente de cada una de las t poblaciones de manera independiente.
2. La variable en estudio debe tener distribución Normal, $X \sim N(\mu_i, \sigma_i)$.
3. Las varianzas de las t poblaciones deben ser iguales, $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2 = \sigma^2$ (homogeneidad de varianzas).

La validez de un experimento depende del muestreo aleatorio y/o del proceso de aleatorización. Para evitar tendencias en los resultados, es esencial que se considere que los datos sean obtenidos de manera aleatoria e independiente de las poblaciones en estudio o que las unidades experimentales sean asignadas aleatoriamente a los t niveles del tratamiento de interés. La falta de Normalidad no afecta seriamente las conclusiones salvo cuando la distribución de la variable es muy asimétrica. La falta de homogeneidad de varianzas afecta el nivel de significación y por lo tanto las conclusiones del análisis.

Para verificar el cumplimiento de estos supuestos se pueden utilizar métodos gráficos y pruebas de hipótesis. Éstos se realizan con un nuevo conjunto de datos el cual es obtenido definiendo una nueva variable denominada residuo cuya expresión es:

$$e_{ij} = X_{ij} - \hat{X}_{ij}$$

donde :

X_{ij}: es el valor obtenido experimentalmente.

$\hat{X}_{ij}$: es el valor estimado por el modelo lineal.

Para más detalles acerca de estos procedimientos se puede consultar Mead, R. et. al. (1993).

Ejercicios de Aplicación:

1.

Con el objetivo de comparar dos dietas de engorde con distintos agregados de levadura (0.6 % y 0.9%) con un testigo sin levadura, se seleccionaron 90 pollos parrilleros machos de 20 días de vida los cuales fueron divididos en tres grupos de 30 y a cada grupo se le asignaron las dietas. Al cabo de cierto tiempo se midió la ganancia de peso para cada pollo, obteniéndose los siguientes valores de ganancia media para cada grupo:

Tratamiento	Media
Levadura al 0.9%	2055.17
Levadura al 0.6%	1952.83
Testigo	1892.17

a) Especificar:
i) la variable en estudio,
ii) la cantidad de poblaciones estadísticas en estudio,
iii) las hipótesis a contrastar.

b) A partir de la siguiente tabla ANOVA extraer conclusiones:

Fuente de Variación	Grados de Libertad	Suma de Cuadrado	Cuadrado Medio	Valor F	Valor p
Tratamiento	2	407215.5556	203607.7778	17.17	**0.0001**
Error	87	1031892.5000	11860.8333		
Total	89	1439108.0556			

c) A partir del siguiente gráfico comentar cuales poblaciones "posiblemente" presentan diferentes medias (recordar que se está trabajando con muestras):

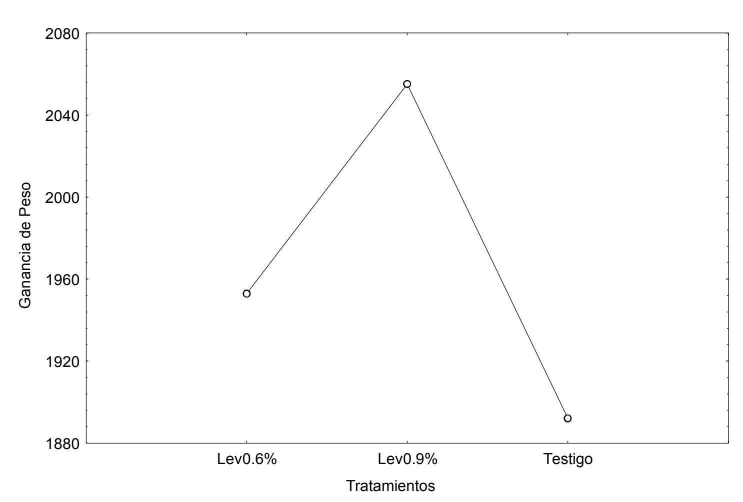


Gráfico de Medias para la variable ganancia de peso

d) La siguiente tabla contiene el valor absoluto de las diferencias de las medias para cada par de tratamientos. El asterisco indica diferencia significativa detectada por la prueba a posteriori de Tukey para un nivel de significación del 5%. Extraer

conclusiones respecto de las medias poblacionales y cotejar con la primera inspección efectuada en el inciso anterior:

Tabla correspondiente al análisis a posteriori de Tukey (α=0.05)

Tratamiento / Tratamiento	Testigo	Levadura al 0.9%
Levadura al 0.6%	60.66	102.34*
Levadura al 0.9%	163*	

2.

Suponga que la experiencia descripta en el ejercicio anterior se hubiese realizado de la siguiente forma: se seleccionan lo 90 pollos parrilleros y se los divide en 15 corrales de 6 aves cada uno, utilizándose 5 corrales por dieta. En cada corral se mide la variable conversión por corral. Se muestra a continuación la tabla de medias correspondientes:

Tratamiento	Media
Levadura al 0.9%	2.2880
Levadura al 0.6%	2.4080
Testigo	2.5160

a) Especificar la unidad experimental y las hipótesis a contrastar.
b) Completar la tabla de análisis de la varianza y extraer conclusiones.

Tabla ANOVA para la variable Conversión

Fuente de Variación	Grados de Libertad	Suma de Cuadrado	Cuadrado Medio	Valor F	Valor p
Tratamiento		0.13008000			
Error		0.02868000			
Total		0.15876000			

c) Concluir a partir de la tabla para el análisis a posteriori de Tukey, respecto de las medias poblacionales.

Tabla correspondiente al análisis a posteriori de Tukey (α=0.05)

Tratamiento / Tratamiento	Testigo	Levadura al 0.9%
Levadura al 0.6%	0.108*	0.12*
Levadura al 0.9%	0.228*	

3.

En un estudio se desea comparar el efecto de tres métodos distintos de congelación de semen porcino para la variable motilidad espermática. La experiencia se realizó de la siguiente manera: cada eyaculado de un cerdo fue dividido en tres partes y cada una de ella fue sometida a un método de congelación: Macropajuela, Micropajuela y Pastilla. Completar la tabla anova que se presenta a continuación y establecer conclusiones

Fuente de Variación	Grados de Libertad	Suma de Cuadrado	Cuadrado Medio	Valor F	Valor p
Tratamiento		10891.85			
Error		17343.85			
Total	119	28235.70			

Tabla correspondiente al análisis a posteriori de Tukey (α=0.05)

Tratamiento / Tratamiento	m	M
P	21.125*	19.15*
M	1.975	

9 Correlación y Regresión Lineal

Objetivos:

- ♦ Distinguir entre un análisis de Regresión Lineal y uno de Correlación Lineal.
- ♦ Interpretar la información obtenida en cada análisis.

9.1 Introducción

En capítulos anteriores se ha trabajado con situaciones que involucran la medición de una única variable sobre cada unidad experimental. Sin embargo, existen una gran variedad de problemas que requieren la consideración simultánea de dos o más variables y el objetivo es estudiar la relación que existe entre ellas. En consecuencia, es necesario estudiar técnicas para analizar problemas de este tipo.

Estas nuevas situaciones llevan a introducir los conceptos de Correlación y Regresión Lineal. Las aplicaciones de estas técnicas son numerosas y se dan en casi todas las ciencias incluyendo ciencias biológicas, ingeniería, física, economía y ciencias sociales entre otras.

En este texto sólo se trata el caso de *Correlación Lineal Simple y Regresión Lineal Simple* es decir, cuando se tienen *dos variables* (denotadas por X e Y) medidas sobre cada unidad experimental y la relación subyacente entre ambas es de tipo *lineal*.

A continuación se muestran algunas situaciones en las cuales el interés es estudiar la relación entre las variables analizadas:

- Un médico desea saber si hay relación lineal entre la presión sanguínea y el nivel de colesterol.
- Un veterinario desea saber si hay relación lineal entre el peso de ciertos animales y su altura.
- Un agrónomo desea predecir el rendimiento de cierta especie en base al contenido de nitrógeno del suelo.
- Un biólogo desea predecir el peso del hígado de ciertos animales en función del peso corporal.

No siempre es claro que tipo de análisis estadístico se debe hacer en situaciones como las planteadas anteriormente por la estrecha relación matemática entre los dos métodos de análisis, pudiéndose pasar fácilmente de uno al otro. Este texto trata de realizar una presentación clara que permita distinguir estos dos conceptos.

9.2 Correlación Lineal Simple

Se llama *Correlación* a la interrelación que existe entre dos variables aleatorias cuantitativas continuas, medidas sobre cada unidad experimental de una misma población.

El estudio de *Correlación Lineal Simple* tiene como objetivo determinar si dos variables están relacionadas o no. Por ejemplo, si a los aumentos de presión sanguínea corresponden aumentos en el nivel de colesterol: en este caso, se dice que hay asociación lineal entre las variables "presión sanguínea" y "nivel de colesterol".

Con la correlación puede investigarse si dos variables X e Y son independientes o si *covarían*, esto es, si varían conjuntamente. Ninguna de estas variables está restringida por el experimentador, o sea que, sobre cada unidad experimental se miden las dos variables[1] (ambas son variables aleatorias).

En el Problema 1.5 se plantea la siguiente situación "Un grupo de investigadores sospecha que hay relación lineal entre el peso y el volumen sanguíneo de cabras de una cierta raza". Los datos, ya presentados en aquel problema, son:

X: Peso (kg.)	34	28	19	41	21	20	21	39	37	23	17	48
Y: Volumen (cm^3)	2.3	2.1	1.1	2.8	1.5	1.6	1.4	2.4	2.5	1.5	1.1	3.5

En el Gráfico 7 del Capítulo 1 se presentó el Diagrama de Dispersión; de acuerdo al comportamiento de la nube de puntos, se puede pensar que hay asociación lineal positiva entre las variables (como se mencionara en aquel capítulo). Otras formas de asociación se presentaron en el Gráfico 8.

La representación gráfica sirve sólo para dar una idea general de la asociación existente entre las variables, pero no alcanza para dar una medida cuantitativa de dicha asociación. Una de estas medidas es el coeficiente de correlación lineal poblacional. Un estimador de éste fue presentado en la Sección 2.5 (Capítulo 2). A continuación se describe el parámetro poblacional correspondiente.

9.2.1 Medida de la Correlación - Coeficiente de Correlación Lineal

Una medida del grado de asociación lineal entre dos variables es la covarianza.

Definición 1: Sean X e Y dos variables aleatorias con distribución normal con $E(X)=\mu_X$, $Var(X)=\sigma_X^2$ y $E(Y)=\mu_Y$, $Var(Y)=\sigma_Y^2$. La covarianza entre las variables X e Y es

$$Cov(X,Y)=E[(X-E(X)(Y-E(Y)]=\sigma_{XY}$$

El inconveniente de esta medida es que su magnitud depende de las unidades empleadas para medir las variables. Por esta razón es necesario estandarizar la covarianza para disponer de una buena medida del ajuste. Esto se obtiene con el coeficiente de correlación lineal.

Definición 2: Sean X e Y dos variables aleatorias con distribución normal con $E(X)=\mu_X$ $Var(X)=\sigma_X^2$ y $E(Y)=\mu_Y$, $Var(Y)=\sigma_Y^2$.El coeficiente de correlación lineal es

$$\rho=\frac{Cov(X,Y)}{\left[E(X-E(X))^2E(Y-E(Y))^2\right]^{1/2}}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

[1] *Notar que (X,Y) son variables aleatorias bidimensionales*.

Este coeficiente es independiente de las unidades de medida, es decir es una cantidad absoluta, sin dimensión.

9.2.1.1 Características del Coeficiente de Correlación

1. Los valores posibles del coeficiente están en el intervalo $[-1,1]$ o sea que $-1 \leq \rho \leq 1$.

2. Cuando el coeficiente toma valores en el intervalo $(0,1]$, indica correlación directa o positiva.

3. Cuando el coeficiente toma valores en el intervalo $[-1,0)$, indica correlación indirecta o negativa.

4. Cuando el coeficiente toma valor cero indica ausencia de correlación.

Cuanto más cerca está ρ de 1 o de -1 mayor es el grado de asociación entre las variables. El coeficiente ρ es un parámetro poblacional y por tanto es un valor constante pero desconocido. Su estimador $\mathbf{r}$, en cambio, es una variable aleatoria dado que depende de los valores muestrales. Una expresión equivalente a la dada en el Capítulo 2 es

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right]^{1/2}}$$

Para el cálculo se puede utilizar la siguiente fórmula de trabajo:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)^{1/2}\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right)^{1/2}}$$

El coeficiente de correlación lineal se denomina “correlación producto-momento de Pearson”.

Para el problema

$$n = 12 \quad \sum_{i=1}^{12} X_i = 348 \quad \sum_{i=1}^{12} X_1^2 = 11276 \quad \overline{X} = 29$$

$$\sum_{i=1}^{12} Y_i = 23.8 \quad \sum_{i=1}^{12} Y_i^2 = 53.24 \quad \overline{Y} = 1.983 \quad \sum_{i=1}^{12} X_i Y_i = 772.9$$

$$r = \frac{772.9 - \frac{1}{12}348 \bullet 23.8}{\left[\left(11276 - 12 \bullet (29)^2\right) \bullet \left(53.24 - 12 \bullet (1.983)^2\right)\right]^{1/2}} = 0.98$$

Este valor es una medida descriptiva para esta muestra particular. Como se puede observar es un valor cercano a 1, lo que indicaría que en esta muestra hay alta correlación positiva. Para poder decidir lo que ocurre en la población se debe realizar una prueba de hipótesis.

9.2.2 Prueba de Significación para el Coeficiente de Correlación

En general se plantean las siguientes hipótesis:

1. $H_0: \rho = 0$ $\qquad$ $H_1: \rho \neq 0$

las cuales indican

H_0: No hay asociación lineal entre las variables X e Y.
H_1: Hay asociación lineal entre las variables X e Y.

2. El estadístico es

$$\varepsilon = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

Si vale H_0, $\varepsilon \sim t_{n-2}$ central.

Si no vale H_0, $\varepsilon \sim t_{n-2}$ no central.

De acuerdo a la hipótesis alternativa la zona de rechazo tiene la forma $Z=(-\infty,-a]\cup[a,+\infty)$ y se completa la prueba de la manera usual.

Si las variables X e Y tienen distribución Normal, probar que éstas son independientes equivale a probar que el coeficiente de correlación ρ es igual a cero. Como esta prueba se realiza bajo el supuesto de normalidad se está probando independencia entre las variables.

A continuación se realiza la prueba de hipótesis para el caso particular del problema planteado.

1. $H_0: \rho = 0$ $\qquad$ $H_1: \rho \neq 0$

las cuales indican

H_0: No hay asociación lineal entre el peso y el volumen sanguíneo de las cabras.
H_1: Hay asociación lineal entre el peso y el volumen sanguíneo de las cabras.

2. El estadístico es

$$\varepsilon = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

Si vale H_0, $\varepsilon \sim t_{10}$ central.

Si no vale H_0, $\varepsilon \sim t_{10}$ no central .

3. Si se fija un nivel de significación $\alpha=0.05$, el valor crítico es $\mathbf{a}=t_{10,0.975}=2.23$ (ver Tabla D del Apéndice) luego la zona de rechazo es $Z=(-\infty,-2.23]\cup[2.23,+\infty)$.

4. El estadístico de contraste calculado es

$$\varepsilon_c = \frac{0.98}{\sqrt{1-(0.98)^2}}\sqrt{12-2} = 15.57$$

En el siguiente gráfico se indica la zona de rechazo y el nivel de significación.

Gráfico 1: Nivel de significación y zona de rechazo
para el Problema 1.5

5. Como $\varepsilon_c = 15.57$ pertenece a la zona de rechazo, la decisión es que se rechaza la hipótesis nula con probabilidad de cometer error de tipo I de 0.05.

6. Conclusión: *hay asociación lineal entre el peso y volumen sanguíneo de las cabras.*

9.3 Regresión Lineal Simple

A veces, cuando se posee información acerca de dos variables cuantitativas, es natural desear expresar una relación funcional entre ellas. El análisis de Regresión es una técnica para investigar y modelar la relación entre variables. En este contexto se estudia la relación entre una *variable aleatoria o no controlada* (Y) y una variable *no aleatoria o controlada* por el investigador (X). El objetivo que será abordado en este texto es el de realizar predicciones, es decir estimar el valor de la variable Y para un valor dado de la variable X. Para realizar este análisis se supone que la relación entre las variables es lineal.

Para desarrollar este tema se plantea el siguiente

Ejemplo 1: A partir de cierto estudio se sabe que la relación entre la presión sanguínea en animales y la dosis de cierta droga es lineal. Se desea predecir la presión para determinadas dosis de la droga. Para ello se consideraron animales de la misma especie, peso y edad a los que se les aplica la droga en diferentes dosis prefijadas por el experimentador y luego se les midió la presión. Los datos se muestran a continuación:

Tabla 2: Dosis de droga y presión sanguínea de animales

Dosis de droga (μg)	2	4	6	8	10
Presión Sanguínea (mm. Hg)	40	75	100	150	180
	42	82	120	130	175
	50	88	110	155	181
	57				

En primer lugar se realiza un diagrama de dispersión.

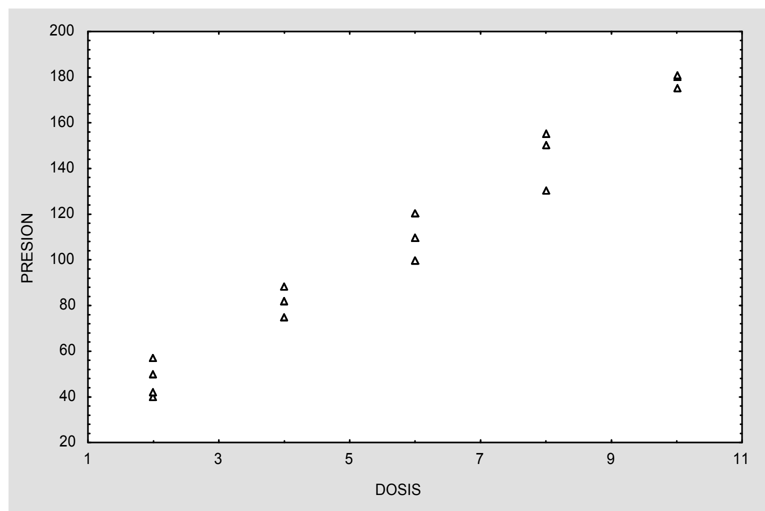


Gráfico 2: Diagrama de dispersión de presión sanguínea y dosis de droga

El diagrama muestra que los datos están (aproximadamente) sobre una línea.

9.3.1 Modelo Lineal

La relación lineal entre las variables X e Y puede expresarse usando el modelo de regresión lineal simple poblacional

$$Y_i = \alpha + \beta X_i + \varepsilon_i \tag{9.1}$$

donde

α: ordenada al origen; β: pendiente; ε_i: componente del error aleatorio.

Los parámetros del modelo son α y β (constantes desconocidas). Como no todos los puntos caen exactamente sobre la recta, se incluye en el modelo el término correspondiente al error aleatorio que es una variable aleatoria con distribución normal con media 0 y varianza σ^2. Por lo tanto las Y_i son variables aleatorias independientes con distribución normal con esperanza $\alpha+\beta X_i$ y varianza común σ^2.(Meyer, P. 1992).

9.3.2 Estimación de los parámetros

Como el modelo (9.1) es poblacional, el primer objetivo es estimar los parámetros. Para ello se debe realizar un experimento y así generar los datos muestrales, en base a los cuales se obtienen los *estimadores* de α y β (denotados por a y b, respectivamente). Luego la recta estimada es

$$\hat{Y}_i = a + bX_i, \qquad i=1,2,...,n$$

Los estimadores $\mathbf{a}$ y $\mathbf{b}$ se denominan *estimadores de mínimos cuadrados*, dado que ellos se obtienen a través del Método de Mínimos Cuadrados. Éste consiste en minimizar las diferencias entre el valor observado Y_i y el valor estimado $\hat{Y}_i$, denominadas residuos (denotados e_i). Geométricamente esto se puede observar en el Gráfico 3, en el cual se

muestran los valores observados (experimentales) y los valores estimados, que están sobre la recta.

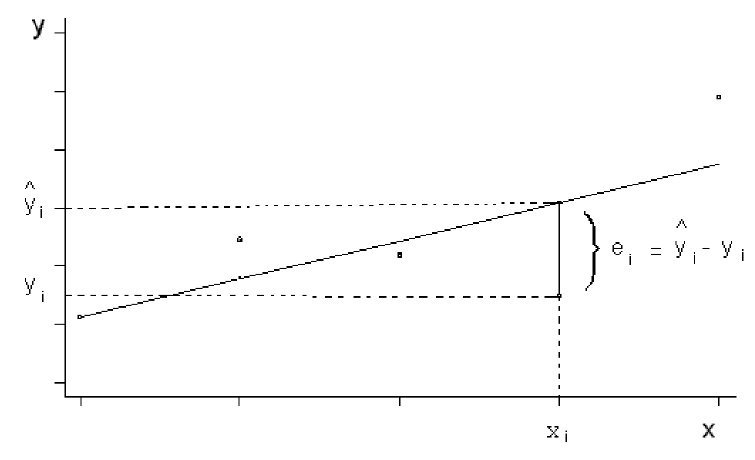


Gráfico 3: Valores observados (Y_i) y valores estimados $\left(\hat{Y}_i\right)$

Minimizando la siguiente expresión

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

se obtienen las fórmulas de los estimadores **a** y **b**, las cuales son

$$b = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} \qquad \text{y} \qquad a = \overline{Y} - b\overline{X}$$

Como **a** y **b** son estimadores mínimos cuadrados la recta estimada ($\hat{Y}_i = a + bX_i$) pasa por el centro de la nube de puntos, y la distancia entre los valores obtenidos experimentalmente (Y_i) y aquellos estimados por esta recta ($\hat{Y}_i$) es mínima. De esta manera se ha construido la recta que tiene la menor distancia con todos los puntos.

El estimador **a** (por ser la ordenada al origen), indica el punto en que la recta corta al eje de las ordenadas, en tanto que **b** (por ser la pendiente de la recta) indica el cambio producido en Y al cambiar en una unidad la X.

Para el Ejemplo 1 se tiene

$$n=16 \qquad \sum_{i=1}^{16} X_i = 92 \qquad \sum_{i=1}^{16} X_i^2 = 664 \qquad \overline{X} = 5.75$$

$$\sum_{i=1}^{16} Y_i = 1735 \qquad \sum_{i=1}^{16} Y_i^2 = 224917 \qquad \overline{Y} = 108.44 \qquad \sum_{i=1}^{16} X_i Y_i = 12178$$

a partir de los cuales los valores de **a** y **b** son

$$b = \frac{12178 - \frac{1}{16} \bullet 92 \bullet 1735}{664 - 16 \bullet (5.75)^2} = 16.31 \qquad a=108.44-5.75\bullet 16.31=14.66$$

luego la recta estimada que resulta es:

$$\hat{Y}_i = 14.66 + 16.31 X_i$$

En este caso el estimador **b**=16.31 mm. Hg/μg indica que *la presión aumenta 16.31 mm. Hg al aumentar en un μg la dosis de droga*. Mientras que **a**=14.66 mm. Hg indica que *cuando el animal no recibe droga (X=0) la presión sanguínea es 14.66 mm. Hg.*

El *gráfico de ajuste del modelo* consiste en bosquejar simultáneamente el diagrama de dispersión de los datos y la recta de regresión estimada. Si los puntos del diagrama están cercanos a la recta hay indicios de que el modelo es adecuado. El análisis no es taxativo ya que es necesario completar con otras técnicas de diagnóstico, que no serán presentadas en este texto (Montgomery, D. y Peck, E. 1982).

Para el Ejemplo 1, el Gráfico 4 es un indicio de que el modelo puede ser el adecuado.

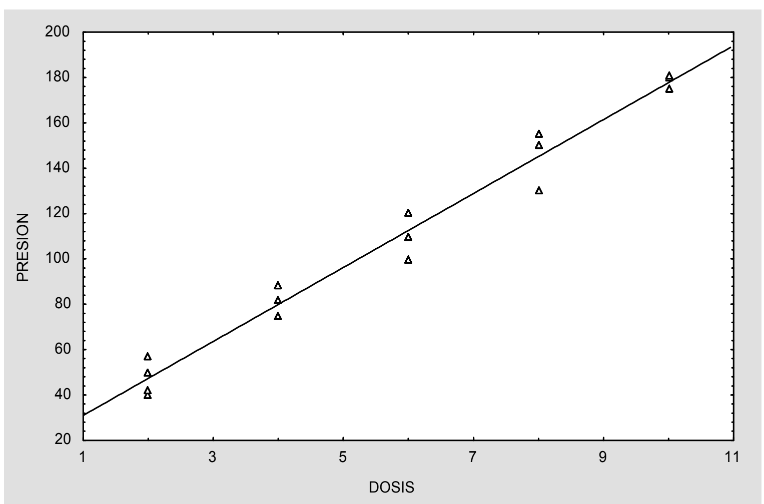


Gráfico 4: Gráfico de ajuste del modelo para
los datos del Ejemplo 1.

9.3.3 Distribución de los Estimadores a y b

Si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes con distribución normal con media $\alpha + \beta X_i$ y varianza σ^2 se puede demostrar que los estimadores **a** y **b** (que son variables aleatorias) tienen asociada las siguientes distribuciones de probabilidades.

$$b \sim N\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right) \qquad a \sim N\left(\alpha, \sigma\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{S_{xx}}}\right)$$

donde $S_{xx} = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$ (Montgomery, D. y Peck, E.1982)

La recta estimada es obtenida a partir de la información de los datos experimentales, luego expresa lo que ocurre en la muestra, es decir la recta estimada es meramente descriptiva. En caso que se desee obtener conclusiones sobre la población para predecir valores de la variable Y para determinados valores de la variable X se deben realizar pruebas de hipótesis.

9.3.4 Pruebas de Significación de los Parámetros

9.3.4.1 Prueba de Significación de la Regresión

Se realiza en primer lugar una prueba para el parámetro β, la cual es llamada *Prueba de Significación de la Regresión*.

Las hipótesis a probar son:

H_0: $\beta = 0$ $\qquad$ H_1: $\beta \neq 0$

que indican

H_0: La variable X no explica linealmente a la variable Y.
H_1: La variable X explica linealmente a la variable Y.

Lo planteado en la hipótesis nula indica que la recta poblacional tiene pendiente cero (o sea es una recta horizontal) lo cual se interpreta como que cualquiera sea la variación en X, Y permanece constante. La estimación está dada por $a = \overline{Y}$.

Para contrastar estas hipótesis se define el siguiente estadístico

$$\varepsilon = \frac{b\sqrt{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}}{s_e}$$

Si vale H_0, $\varepsilon \sim t_{n-2}$ central.

Si no vale H_0, $\varepsilon \sim t_{n-2}$ no central

donde $s_e = \left[\frac{\sum_{i=1}^{n} Y_i^2 - a\sum_{i=1}^{n} Y_i - b\sum_{i=1}^{n} X_i Y_i}{n-2}\right]^{\frac{1}{2}}$

La *zona de rechazo* es $Z=(-\infty,-a]\cup[a,+\infty)$ donde $\mathbf{a}=t_{n-2,\ 1-\alpha/2}$ (de la Tabla D del Apéndice).

Si se rechaza H_0 la variable X explica linealmente a la variable Y.

En caso de no rechazar H_0 no significa necesariamente que las variables X e Y no están relacionadas linealmente. El hecho de no poder mostrar estadísticamente que la pendiente es distinta de cero, puede deberse a una gran variabilidad en los datos producida por el proceso de medición de los mismos o a porque el rango de los valores de la variable X no es el apropiado. Para concluir que $\beta=0$ se requiere una gran variedad de evidencia no estadística y conocimiento del problema.

En muchas situaciones prácticas la variable aleatoria Y es explicada linealmente por más de una variable controlada. Cuando esto sucede, se dice que el modelo adecuado es un modelo de Regresión Lineal Múltiple. El análisis que se realiza es análogo al presentado, pero un tanto más complejo (Montgomery, D. y Peck, E.1982)

Los resultados de la prueba son válidos si la variable aleatoria Y tiene distribución normal y el modelo (9.1) es el adecuado.

Retomando el Ejemplo 1, se tiene

1. H_0: $\beta=0$ H_1: $\beta\neq 0$

que indican

H_0: La dosis de droga no explica linealmente a la presión sanguínea

H_1: La dosis de droga explica linealmente a la presión sanguínea

2. En base a los datos, $\varepsilon_c=24.052$.

3. Si el nivel de significación elegido es $\alpha=0.05$ el valor de critico es $\mathbf{a} = t_{14;0.975}=2.14$, luego $Z = (-\infty;-2.14]\cup[2.14;+\infty)$. Gráficamente

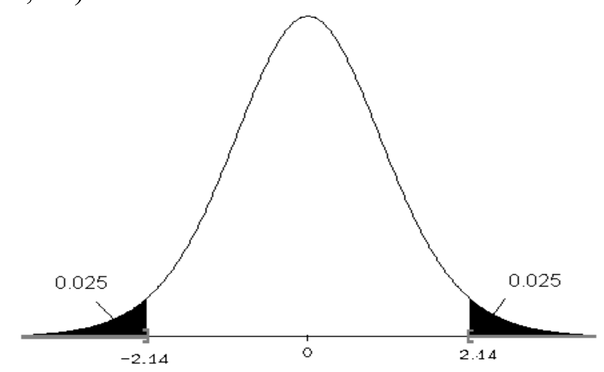


Gráfico 5: Nivel de significación y zona de rechazo

4. Como el estadístico contraste 24.052 pertenece a la zona de rechazo Z, la *decisión* es rechazar la hipótesis nula con probabilidad de cometer error de tipo I de 0.05.

5. La conclusión es: *la dosis de la droga explica linealmente a la presión sanguínea*.

Si el estadístico de contraste no perteneciera a la zona de rechazo, la decisión sería no rechazar la hipótesis nula con probabilidad de cometer error de tipo II, en cuyo caso la conclusión sería: *no hay evidencias de que la dosis de la droga explique linealmente a la presión sanguínea*.

9.3.4.2 Prueba de Significación de la Ordenada al Origen

Esta prueba debería realizarse sólo cuando la regresión es significativa, o sea cuando se rechaza la hipótesis nula en la prueba anterior.

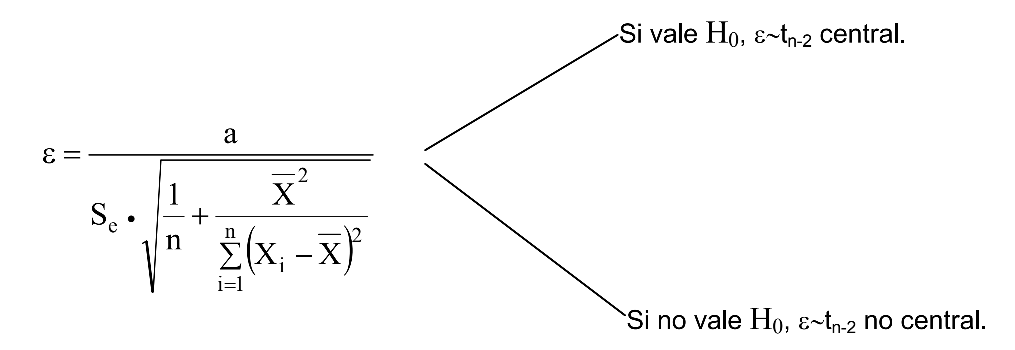
Las hipótesis son

1. $H_0:\alpha=0$ $H_1:\alpha\neq 0$

que indican

H_0: La recta pasa por el origen.
H_1: La recta no pasa por el origen.

Para contrastar estas hipótesis se define el siguiente estadístico



$$\varepsilon = \frac{a}{S_e \bullet \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}}}$$

y donde S_e está definido como antes.

La zona de rechazo es $Z=(-\infty;-a]\cup[a;+\infty)$, que se determina de la forma acostumbrada.

Si se rechaza H_0, se puede afirmar que la recta no pasa por el origen.

En caso de no rechazar H_0, se dice que no hay evidencia para afirmar que recta no pasa por el origen. Esto dice que podría establecerse entre las variables X e Y un modelo sin ordenada como el siguiente

$$Y=\beta X$$

con lo cual se debería comenzar de nuevo el análisis con este modelo en vez del propuesto en (9.1). Cabe señalar que este modelo plantea que al valor 0 de la variable X corresponde el valor 0 de la variable Y, lo cual debería tener sentido en el contexto del problema.

Para el Ejemplo 1 la prueba de significación de la ordenada al origen es:

1. $H_0:\alpha=0$ $\qquad$ $H_1:\alpha\neq 0$

que indican

H_0: La recta pasa por el origen.
H_1: La recta no pasa por el origen.

2. En base a los datos $\varepsilon_c= 3.356$.

3. Si $\alpha=0.05$, el valor de critico es $\mathbf{a}= t_{14;0.975}=2.14$ (Tabla D del Apéndice). Luego $Z=(-\infty;-2.14]\cup[2.14;+\infty)$. Esta zona coincide con la presentada en el Gráfico 5.

4. Como el estadístico de contraste pertenece a la zona de rechazo, la decisión es rechazar la hipótesis nula con probabilidad de cometer error de tipo I de 0.05.

5. La conclusión es: *la recta no pasa por el origen*.

9.3.5 Utilidad de la recta de regresión estimada

La recta de regresión estimada se puede utilizar para predecir el valor de Y para un X determinado, por ejemplo para X_0. Para que esta predicción sea posible se deben verificar las siguientes condiciones

- Se debe rechazar H_0 en las pruebas de significación de la regresión y de la ordenada al origen.
- El valor de $X=X_0$ debe estar entre los valores mínimo y máximo, esto es $X_{min} \leq X_0 \leq X_{max}$.

La segunda condición expresa que la relación establecida por la recta es válida sólo en el rango de los X considerados. Fuera de este rango la relación entre las variables puede no ser la propuesta en (9.1).

Para el Ejemplo 1 se desea estimar la presión sanguínea para una dosis de droga de 7 μg. Dado que se rechazó H_0 en ambas pruebas de hipótesis y que $X_0=7$ está entre el valor mínimo y máximo, reemplazando en la recta estimada se obtiene el valor deseado, que indica que para una dosis de 7 μg se estima una presión de 128.83 mm. Hg.

9.3.6 Coeficiente de Determinación

El valor

$$R^2 = \frac{b\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i Y_j - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{j=1}^{n} Y_j\right)}{\sum_{j=1}^{n} Y_j^2 - n\overline{Y}^2}$$

es llamado *coeficiente de determinación* e indica la *proporción de variación de Y explicada por la variable regresora X.*

Para el Ejemplo 1, R^2=0.976 el cual indica que 97.6*% de las variaciones producidas en la presión sanguínea están explicadas por la dosis de droga*.

9.4 Consideraciones finales para el uso de la Correlación y la Regresión

1. Para poder realizar inferencias, tanto en el Análisis de Regresión como en el de Correlación, se debe verificar el cumplimiento de ciertos supuestos (Montgomery, D. y Peck, E.1982)
2. Cuando se realiza el Análisis de Correlación Lineal Simple se supone (por razones no estadísticas) que tiene sentido asociar linealmente las variables, aunque el objetivo no es encontrar la forma de esa relación.
3. Cuando se realiza un Análisis de Regresión Lineal Simple las conclusiones de las pruebas de significación de los parámetros son válidas si el modelo (9.1) es el más adecuado para

describir la relación entre las variables X e Y. Para determinar si tal modelo es el adecuado se debe realizar un análisis que no está al alcance de este libro (Montgomery, D. y Peck, E.1982). En los ejercicios de aplicación se muestran algunas técnicas de diagnóstico sencillas.

4. La utilización de una u otra técnica depende fundamentalmente del objetivo del investigador. Una vez elegida, la misma técnica condiciona la manera de realizar la experiencia para que sean válidos los resultados provistos por ella. Si se realiza un Análisis de Correlación, para llevar a cabo la experiencia se miden (sobre cada unidad experimental) las dos variables de interés; en cambio, cuando se lleva a cabo un análisis de Regresión, sobre cada unidad experimental se mide una de las variables (Y) habiendo fijado de antemano los valores de la otra variable (X).

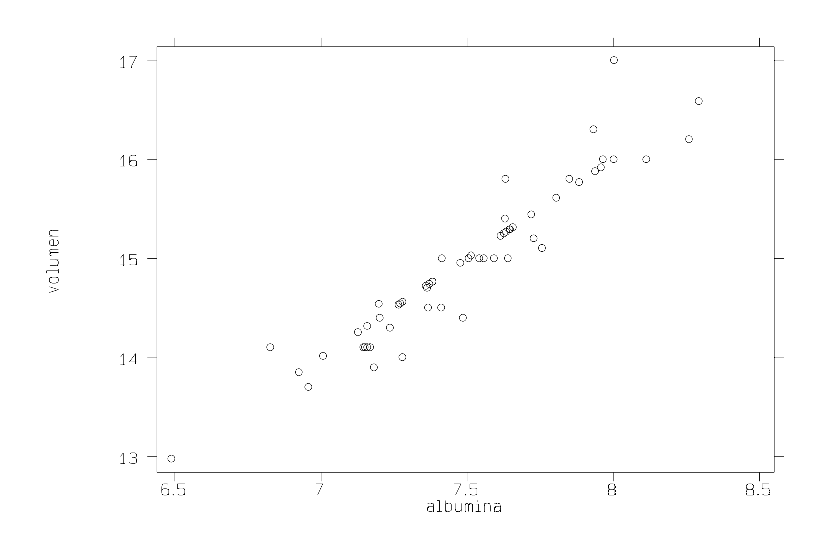
5. Para completar la información provista por ambos análisis, se podrían construir intervalos de confianza para los parámetros respectivos aunque no serán mostrados en este texto.

Ejercicios de Aplicación

1.

Para estudiar si el volumen de plasma (X) y la albúmina circulante (Y) están asociados linealmente se seleccionaron al azar 58 varones, a los que se les midieron las dos variables.

a) Discutir la presencia o ausencia de asociación lineal que podría sugerir el siguiente diagrama de dispersión para los datos de volumen y albúmina.



b) Enunciar las hipótesis correspondientes a la prueba de correlación.

c) El valor de **r** para los datos de este problema es 0.9509, ¿Concuerda este valor con las observaciones que ha realizado en el inciso a)?

d) El valor **p** para la prueba de hipótesis de correlación es cero. Extraer conclusiones respecto del problema en cuestión.

2.

En cierta investigación forestal se ha planteado como objetivo estudiar la posible asociación lineal entre el crecimiento en altura de los árboles y el aumento de su

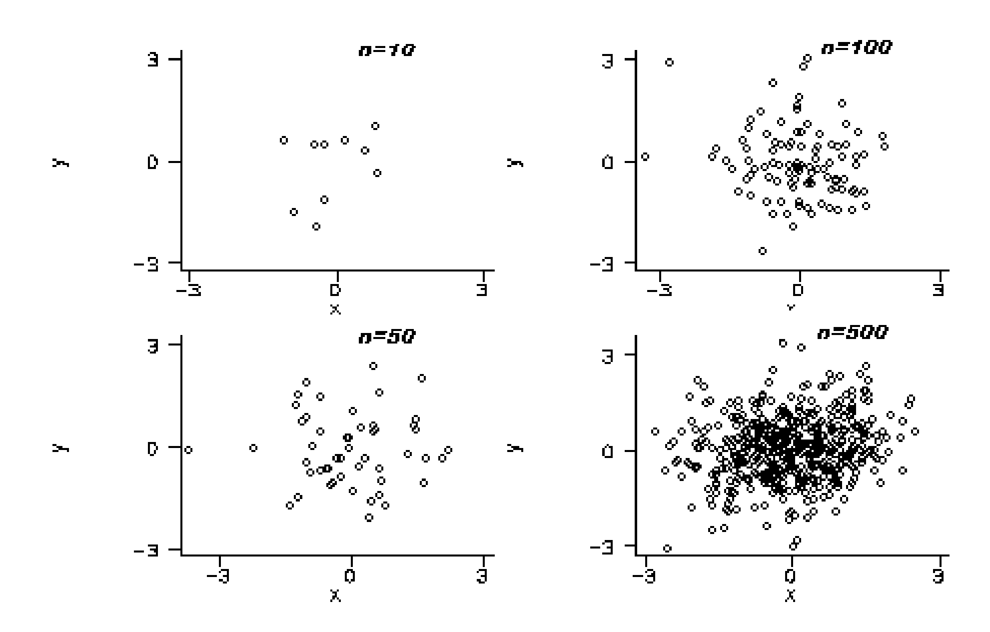
diámetro, seleccionándose para tal fin 9 árboles al azar a los que se les midieron ambas variables. Los datos obtenidos se presentan en la siguiente tabla:

Diámetro (metros)	0.2	0.4	0.5	0.5	0.6	0.65	0.65	0.7	0.7
Altura (metros)	4.3	4.4	5.3	6.3	6.4	6.5	7.8	8.3	8.5

Realizar un análisis completo de correlación incluyendo el diagrama de dispersión.

3.

En el siguiente gráfico se muestran cuatro diagramas de dispersión correspondientes a muestras de diferentes tamaños pero con el mismo valor del coeficiente de correlación muestral.



a) A partir de la inspección de los diagramas, ¿ el valor **r** es alto o bajo?.

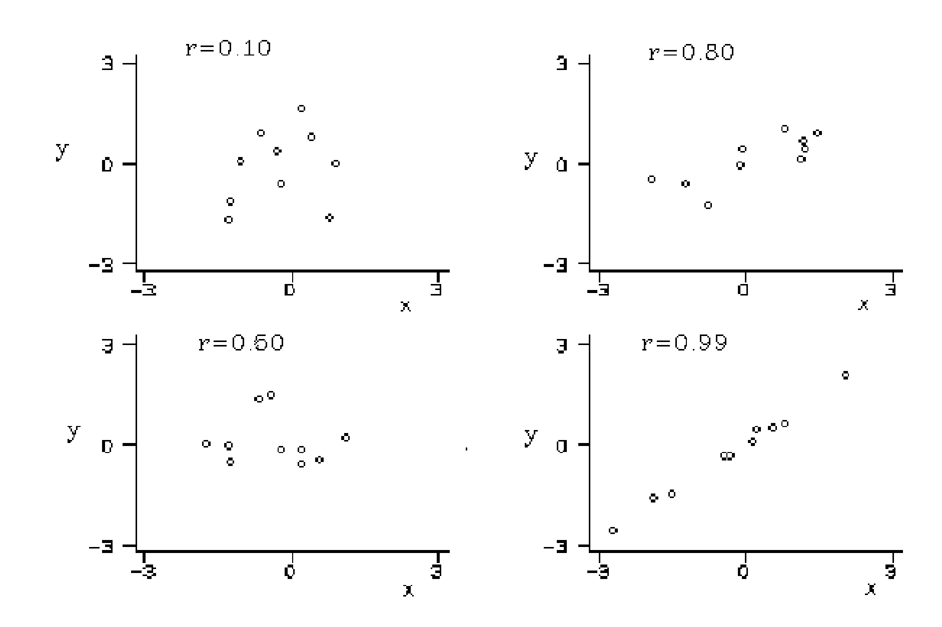
b) Completar la siguiente tabla, suponiendo que el valor de **r** para los diagramas anteriores es 0.1:

Tamaño de Muestra	Valor del Estadístico Calculado para la prueba de correlación	Valor Crítico de la Prueba	Decisión respecto de H_0 : ρ=0
n=10			
n=50			
n=100			
n=500			

c) A partir de la tabla anterior, discutir la relación entre la sensibilidad de la prueba de correlación y el tamaño muestral.

4.

En el siguiente gráfico se muestran cuatro diagramas de dispersión correspondientes a muestras del mismo tamaño (n=10) pero con distintos valores del coeficiente de correlación muestral.



Realizar un análisis similar al efectuado en el ejercicio anterior, teniendo en cuenta que en esta nueva situación el objetivo es estudiar la relación entre coeficiente de correlación muestral y sensibilidad de la prueba de hipótesis.

5.

Con el objetivo de obtener predicciones del número de bacterias en función del tiempo, se midió el número de bacterias en 7 tiempos seleccionados previamente, obteniéndose los siguientes datos:

Tiempo (hs.)	0	1	2	3	4	5	6
Número de bacterias	20	57	75	102	142	190	200

a) ¿Cuáles son las variables en estudio? ¿Cuál debería elegir como variable X y cuál como variable Y? Justificar.

b) Realizar el diagrama de dispersión.

c) Escribir las hipótesis a contrastar (para ambas pruebas mencionadas).

d) En la siguiente tabla se dan los parámetros estimados y los valores **p** para las pruebas de significación de la regresión y ordenada al origen. Dar la ecuación de la recta de regresión estimada y graficar. Indicar las conclusiones para un nivel de significación del 5%.

Parámetro	Parámetro Estimado	Valor p
β	31.11	0.000
α	18.68	0.038

e) ¿Para cuáles de los valores de tiempo que se indican a continuación es posible predecir el número de bacterias?: x=1.2, x=2, x=3.7. Para aquellos en que sea factible dar el valor de la predicción.

6.

A los efectos de predecir la temperatura del conejo después de haber sido inoculado con virus de morriña en tiempos determinados, se realizó el siguiente experimento: se

seleccionaron al azar 7 conejos y se les inoculó el virus en distintos tiempos prefijados; obteniéndose los siguientes datos:

Tiempo después de la inyección (hs)	Temperatura (ºF)
24	107.3
32	104.5
48	105.5
56	106.0
72	103.9
80	103.2
96	102.1

a) ¿Cuál de las dos variables es posible escoger como regresora y cuál como respuesta?

b) Realizar el diagrama de dispersión para los datos. ¿Sugiere éste que un modelo de regresión lineal es el adecuado?

c) Establecer conclusiones para ambas pruebas de significación a partir de la siguiente información:

Parámetro	Parámetro Estimado	Valor p
β	-0.06	0.01
α	108.04	0.00

Comparar los resultados con el análisis realizado en el inciso ***b)***.

d) A partir del diagrama de dispersión realizar el gráfico de ajuste del modelo y establecer conclusiones.

7.

Un ingeniero agrónomo desea predecir el rendimiento de un cierto forraje a partir de la cantidad de agua que recibe. Para ello sembró dicho forraje en 6 parcelas regadas con distintas cantidades de agua, medidas en cm^3. Las cantidades de agua y los rendimientos obtenidos se indican en la siguiente tabla.

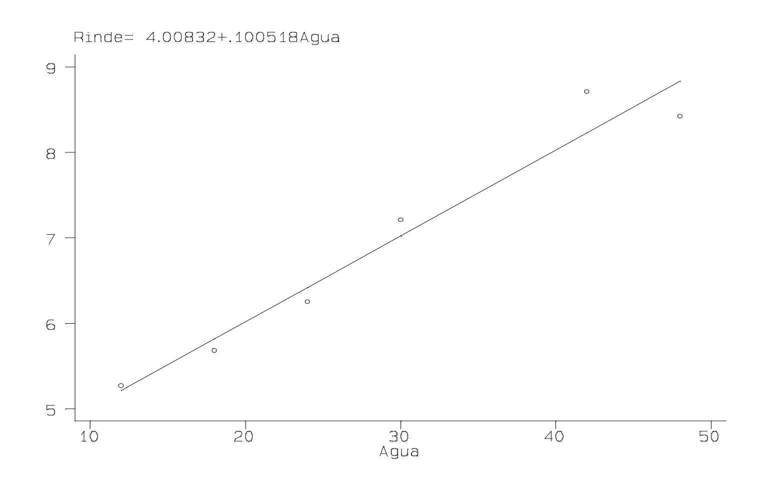
Agua (cm^3)	12	18	24	30	42	48
Rendimiento (tn)	5.27	5.68	6.25	7.21	8.71	8.42

a) ¿Cuál de las variable se debe tomar como regresora y cuál como respuesta?

b) partir de la siguiente tabla concluir respecto del problema que se está abordando.

Parámetro	Parámetro Estimado	Valor p
β	0.1	0.001
α	4.01	0.000

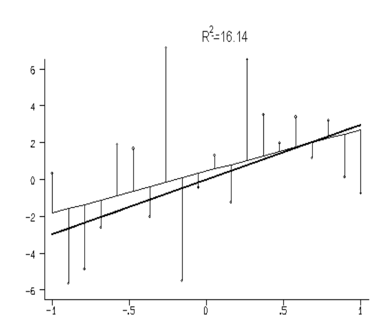
c) A continuación se da el gráfico de ajuste del modelo. Discutir respecto de la adecuación del modelo en cuestión.

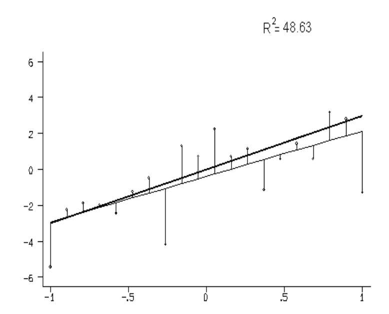


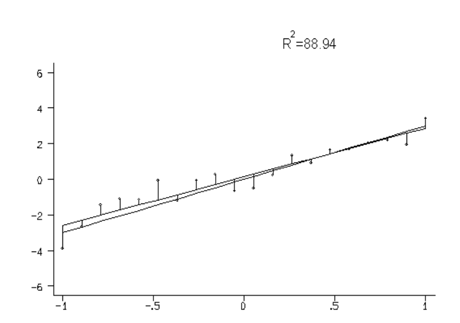
d) Predecir, si es posible, el rinde para 35 cm^3 de agua.

8.

En cada gráfico presentado a continuación se muestran: el diagrama de dispersión, la recta de regresión estimada (en trazo fino), las distancias e_i de cada punto a la recta (marcadas con una línea vertical), la recta de regresión poblacional (en trazo grueso) y el valor de R^2 (multiplicado por 100).







a) Discutir la relación entre el valor R^2 y el ajuste del diagrama a la recta de regresión estimada.

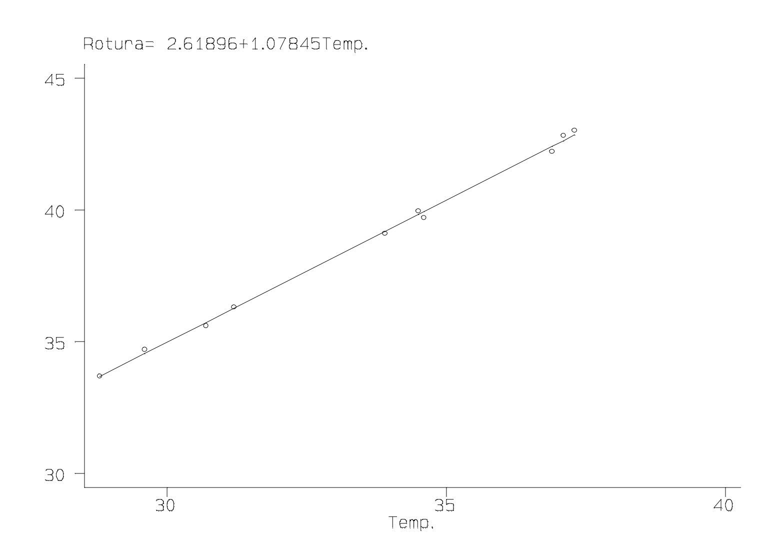
b) ¿A medida que el valor R^2 aumenta, que sucede con las rectas (estimada y poblacional)?.

9.

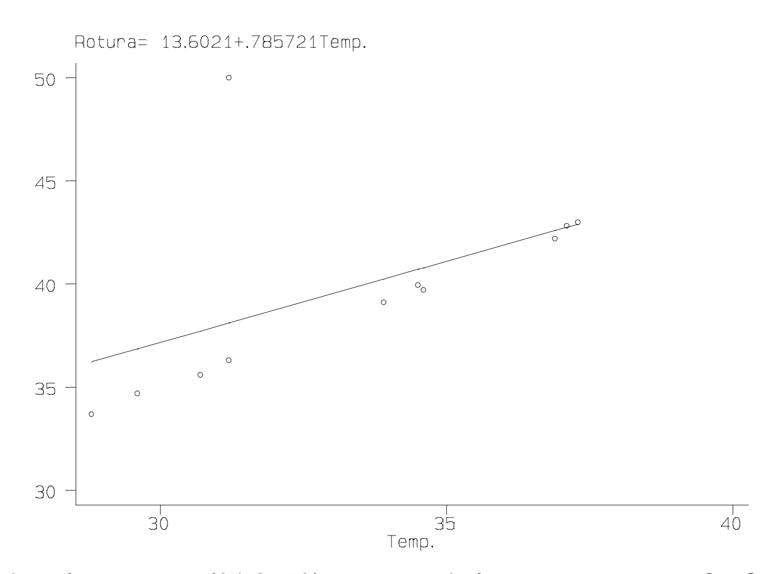
Se dispone de 10 observaciones de temperatura (X) (previamente determinadas) de granos de arroz sin descascarar y sus correspondientes porcentajes de rotura (Y) en la molienda. El objetivo es predecir el porcentaje de rotura para ciertas temperaturas.

Temp.	33.9	34.6	34.5	36.9	37.1	37.3	28.8	29.6	30.7	31.2
Rotura	39.1	39.7	39.95	42.2	42.81	43	33.68	34.7	35.6	36.3

a) El valor **p** para la prueba de significación de la regresión es 0 y el correspondiente a la prueba para la ordenada al origen es 0.002. ¿Qué conclusiones se pueden obtener a partir de estos valores de **p** y del gráfico de ajuste del modelo que se presenta a continuación?

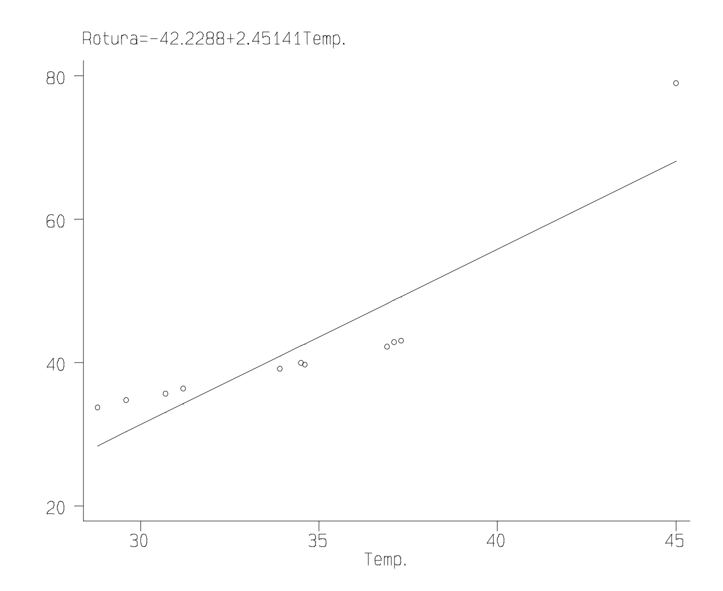


b) Suponga que se efectuó una nueva medición para una temperatura de 31.2 grados y se obtuvo un porcentaje de rotura del 50%. Cuando este dato se incorpora al análisis los valores de **p** son 0.104 y 0.372 para las pruebas relativas a los parámetros β y α, respectivamente. El gráfico de ajuste del modelo es el siguiente:



¿Cómo se ubica el nuevo par (31.2, 50) respecto de los restantes pares? ¿Cuál ha sido el efecto de dicho punto sobre el ajuste del modelo?

c) Si en lugar de incorporar la medición (31.2,50) se hubiese incorporado la medición (45, 79) los valores **p** para las pruebas relativas a los parámetros β y α serían 0 y 0.014 respectivamente. El gráfico de ajuste del modelo es:



Discutir las diferencias entre este caso y el dado en el inciso ***b)***.

10 Prueba de Ji-Cuadrado

Objetivos:

- Distinguir las distintas aplicaciones de la prueba de Ji-Cuadrado.
- Aplicar la prueba a diferentes situaciones.

10.1 Introducción

En este capítulo se analizan algunas situaciones que no se pueden resolver con las técnicas presentadas anteriormente. En capítulos anteriores se analizaron variables de tipo cuantitativo; a continuación se presentan situaciones que involucran además variables cualitativas o categorizadas, que pueden ser analizadas a través de una prueba de hipótesis particular, denominada *Prueba de Ji- Cuadrado*.

Se debe recordar que las *variables cualitativas* son aquellas cuyos diferentes estados se pueden expresar por medio de una cualidad, por ejemplo: sexo, raza, color, nivel educacional, preferencia por un producto, aptitud hacia la lectura. Todas ellas requieren un análisis estadístico diferente a los desarrollados hasta el momento.

A continuación se presentan situaciones en las cuales se puede aplicar la Prueba de Ji-Cuadrado:

1. En el caso de variables cualitativas, se puede estar interesado en saber si sus categorías se manifiestan en una determinada proporción o si las variables están relacionadas. Por ejemplo:

   * Al tirar un dado se desea saber si está equilibrado.

   * En un estudio genético se quiere confirmar si una característica (por ejemplo el color de cabello) se da en una determinada proporción.

   * En encuestas de opinión el objetivo es confirmar la proporción de votos favorables y desfavorables para un candidato.

   * En una experiencia con niños de 12 años se desea determinar si existe relación entre el estado nutricional y el coeficiente intelectual.

2. En el caso de variables cuantitativas (discretas o continuas), se desea establecer si se ajustan a una distribución teórica determinada. Por ejemplo:

   * En cierto estudio se desea determinar si la altura de los adultos se distribuye como una Normal con media 1.71 m. y desvío 0.5 m.

   * Se está interesado en determinar si el Número de animales sanos de una cierta raza se distribuye como una Binomial con parámetros $n=10$ y $p=0.8$.

Para desarrollar las distintas aplicaciones de esta prueba se formulan diferentes situaciones, cada uno de las cuales corresponde a una aplicación distinta. A continuación se

presentan cada una de ellas.

10. 2 Prueba de Concordancia

Ejemplo 1: Un médico tiene la sospecha que cierta enfermedad (que requiere hospitalización) afecta igualmente a hombres y a mujeres, es decir, que se presenta en la misma proporción para los dos sexos.

Sean $\mathbf{p_1}$ y $\mathbf{p_2}$ las proporciones de hombres enfermos y mujeres enfermas, respectivamente. Como sólo hay dos categorías, al decir que se espera la misma proporción, cada una de ellas debería ser ½. Estadísticamente esto se traduce en las siguientes hipótesis

H_0: $p_1=1/2$, $p_2=1/2$ H_1: al menos una distinta

que indican

H_0: Cierta afección para hombres y mujeres se da en la proporción 1:1.
H_1: Lo contrario.

A continuación, para cotejar (confirmar o descartar) esta afirmación se debe realizar un experimento aleatorio. Éste consiste en seleccionar una persona enferma hospitalizada y registrar su sexo, y repetir esto una cierta cantidad de veces prefijada.

Suponga que se seleccionaron 900 pacientes y se los clasificó de acuerdo al sexo (la variable en estudio) y se contó la cantidad de personas enfermas de cada sexo. Dicha información se puede resumir en una tabla de frecuencias como la siguiente:

Tabla 1: Distribución de los enfermos según el sexo

SEXO	f_o
Masculino	480
Femenino	420
TOTAL	900

En el problema planteado se tiene una variable cualitativa *Sexo*, la cual tiene dos categorías (Hombre, Mujer). Para poder trabajar este tipo de variables numéricamente lo que se hace es contar el N° de observaciones en cada categoría de la variable (*frecuencia observada* de cada categoría), con lo que se pasa de una variable aleatoria cualitativa a una variable aleatoria discreta.

Los valores observados en este experimento son:

f_{o1}: n° de pacientes hombres (de un total de 900).
f_{o2}: n° de pacientes mujeres (de un total de 900).

Si se cumpliera la afirmación del médico, de los 900 pacientes enfermos deberían ser 450 hombres y 450 mujeres. Estos números son llamados *frecuencias esperadas* o *teóricas*, las cuales son denotadas generalmente por f_e.

Para saber si la proporción de hombres y mujeres es la misma, es natural basarse en la comparación entre las dos frecuencias (las obtenidas experimentalmente y las propuestas por el experimentador). En este caso, se desea comparar las *frecuencias observadas* (480 hombres y 420 mujeres) con las *frecuencias esperadas* de acuerdo a la afirmación del médico (450 hombres y 450 mujeres).

Parece natural entonces definir un estadístico, para probar estas hipótesis, en función de las frecuencias *esperadas o teóricas* (parámetro a estimar) y de las *frecuencias observadas* (variables aleatorias) (Mendenhall, W. et. al. 1994) Luego el estadístico definido para esta prueba es:

$$\varepsilon = \sum_{i=1}^{k} \frac{(f_{oi} - f_{ei})^2}{f_{ei}}$$

Si vale H_0, $\varepsilon \sim \chi^2_{k-1}$ central

Si no vale H_0, $\varepsilon \sim \chi^2_{k-1}$ no central

donde χ^2_{k-1} indica la distribución Ji-Cuadrado con k grados de libertad (k es la cantidad de categorías de la variable). En el problema planteado se tienen dos categorías hombre-mujer entonces $\varepsilon \sim \chi^2_1$ si vale H_0.

Se puede observar que este estadístico toma siempre valores mayores o iguales a cero pues está definido como un suma de cuadrados dividido un número positivo.

Si la frecuencia teórica y la frecuencia observada son iguales, la diferencia ($f_{oi}-f_{ei}$) para cada categoría es cero y por lo tanto el estadístico es cero, con lo cual no habría dudas acerca de la decisión a tomar: no se puede rechazar la H_0.

Sin embargo, rara vez las frecuencias observadas son exactamente iguales a las esperadas por lo cual el estadístico, en general, resulta un valor diferente de cero. Se debe decidir si esto se debe al azar o a que no se da la proporción esperada.

Para tomar la decisión, se debe determinar el valor crítico **a**, que indicará cuándo las diferencias son lo suficientemente grandes como para considerar f_{oi} diferente de f_{ei}. Luego la zona de rechazo Z de esta prueba es *siempre unilateral derecha*, es decir $Z=[a,+\infty)$.

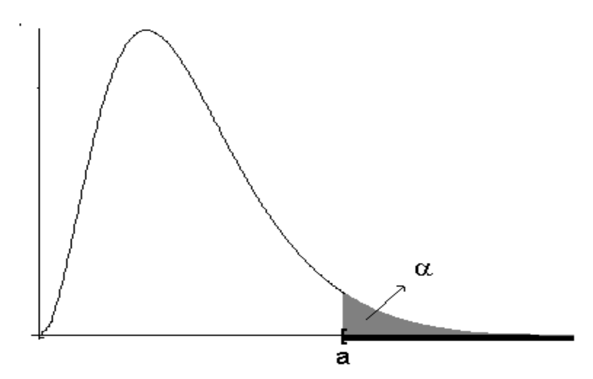


Gráfico 1: Nivel de significación y zona de rechazo

Si en el problema, el nivel de significación considerado para probar las hipótesis es $\alpha=0.01$, entonces el valor crítico es $a = \chi^2_{1,0.01} = 6.63$ (Tabla E del Apéndice).

Para calcular el valor numérico del estadístico se deben determinar las frecuencias esperadas o teóricas f_e, para lo cual se debe tener en cuenta la siguiente definición:

Definición 1: Sea ***E*** un experimento aleatorio y ***S*** el espacio muestral asociado a él. Para un suceso cualquiera ***A*** de ***S***

$$f_{eA}=N \bullet P(A)$$

donde N indica el tamaño de la muestra.

Utilizando esta definición en el Ejemplo 1 se tiene que

$$f_{e1}=900 \bullet \frac{1}{2}=450 \qquad \text{y} \qquad f_{e2}=900 \bullet \frac{1}{2}=450$$

Estos valores coinciden con los que se obtuvieron anteriormente de manera intuitiva, aunque no en todos los casos las frecuencias esperadas pueden ser determinadas en forma inmediata.

El valor numérico del estadístico es:

$$\varepsilon_c=\frac{(480-450)^2}{450}+\frac{(420-450)^2}{450}=\frac{900}{450}+\frac{900}{450}=2+2=4$$

Luego dado que el $\varepsilon_c=4$ y este valor no pertenece a Z, entonces no se rechaza H_0 con probabilidad de cometer error de tipo II y se puede concluir que *no hay evidencia para decir que la enfermedad no se da en la misma proporción en hombres y mujeres* (no hay suficiente evidencia para contradecir la afirmación del médico).

A continuación se detallan algunos aspectos generales de esta prueba.

1. Hipótesis

H_0: $p_1=p_{1.0}$, $p_2=p_{2.0}$,...,$p_k=p_{k.0}$ $\qquad$ H_1: al menos una diferente

donde $p_{1.0}$, $p_{2.0}$,...,$p_{k.0}$ son valores conocidos, no necesariamente iguales

2. Estadístico de la prueba (propuesto por Karl Pearson): es una variable aleatoria pues f_o es variable aleatoria (su valor depende de la muestra). Este estadístico tiene una distribución aproximada Ji-Cuadrado, con $\mathbf{k-1}$ grados de libertad, o sea número de categorías de la variable en estudio menos 1. La prueba matemática correspondiente para determinar que el estadístico asume una distribución aproximadamente Ji-Cuadrado está fuera del alcance de este libro (Agresti, A. 1990).

3. Zona de Rechazo: Para esta prueba la zona de rechazo que corresponde siempre es $Z=[a,+\infty)$ Esta forma particular de la zona de rechazo depende del estadístico como se puede observar a continuación: si las f_o son muy parecidas a las f_e, la decisión correcta es la de no rechazar la hipótesis nula, y en ese caso el valor numérico del estadístico será un valor pequeño (cerca de cero); si las f_o son muy distintas de las f_e el valor numérico del estadístico será grande (lejos de cero) en cuyo caso se debería tomar la decisión de rechazar la hipótesis nula.

Ejemplo 2: Se desea estudiar si un rasgo particular de la mandíbula se considera heredado en la proporción 1:2:1 para homocigota dominante, heterocigota y homocigota recesivo (AA,Aa,aa) respectivamente. Para ello se seleccionan 150 niños al azar, cuyas frecuencias observadas en cada categoría se indican en la siguiente tabla.

Tabla 2: Frecuencias observadas por categorías

Categorías	Dominante	Heterocigota	Recesivo
f_o	31	92	27

Las hipótesis a probar

1. H_0: $p_1=p_{1.0}=1/4$ $p_2=p_{2.0}=1/2$ $p_3=p_{3.0}=1/4$
   H_1: al menos una diferente

es decir

H_0: El rasgo se hereda en la proporción 1:2:1 para Homocigota Dominante (D), Heterocigota (H) y Homocigota Recesivo (R) respectivamente.
H_1: Lo contrario.

2.

$$\varepsilon = \sum_{i=1}^{k} \frac{(f_{oi} - f_{ei})^2}{f_{ei}}$$

Si vale $H_0, \varepsilon \sim \chi^2_{k-1}$ central.

Si no vale $H_0, \varepsilon \sim \chi^2_{k-1}$ no central.

3. Si $\alpha=0.05$, se tiene que el valor crítico es $a = \chi^2_{2,0.05} = 5.99$ (Tabla E).

4. Para calcular el estadístico se necesitan de las frecuencias observadas, que se determinan en forma experimental y de las frecuencias teóricas o esperadas de cada categoría que se determinan utilizando la Definición 1. Para este ejemplo las f_e resultan

$$f_{eD} = n \bullet P(D) = 150 \bullet \frac{1}{4} = 37.50$$

$$f_{eH} = n \bullet P(H) = 150 \bullet \frac{1}{2} = 75$$

$$f_{eR} = n \bullet P(R) = 150 \bullet \frac{1}{4} = 37.50$$

Luego $\varepsilon_c=7.92$.

5. Dado que $7.92 \in Z$, la decisión es rechazar H_0 con P(cometer error de tipo I)$=0.05$

6. Conclusión: *el rasgo particular de la mandíbula no es heredado en la proporción 1:2:1 para Homocigota Dominante, Heterocigota y Homocigota Recesivo*.

10.3 Tablas de Contingencia

Cuando se tienen dos variables de tipo cualitativo y se desea estudiar si existe relación entre ambas o determinar si la proporción de ocurrencia de una de las categorías de una de las variables (éxito) es la misma para todas las categorías de la otra variable. Las frecuencias observadas son arregladas en tablas de contingencias o tablas de doble entrada, que fueron definidas en el Capítulo 1.

Cuando los datos se disponen en este tipo de tablas se pueden presentar dos

situaciones experimentales como las siguientes:

1. Sólo el total general está fijo.
2. Los totales por fila o por columna están fijos.

Cuando la situación experimental es como la expuesta en el caso 1 se realiza una Prueba de Independencia, mientras que para el caso 2 corresponde la prueba de Homogeneidad de Proporciones. Si todos los totales están fijos, corresponde otra prueba que no será presentada en este texto (Ato García, M. y López García, J. 1996).

10.3.1. Prueba de Independencia

Ejemplo 3: Sobre la base de los datos del Problema 1.6, el investigador está interesado en estudiar si existe relación entre el color de pelaje y la presencia de tumores cutáneos en equinos.

Este problema se diferencia de los anteriores por el hecho de que se estudian dos variables de tipo cualitativo: *Color de pelaje* y *Presencia de tumor*. Cada variable está dividida en categorías excluyentes: la variable Color tiene tres categorías (Alazán, Zaino, Tordillo) y la variable Presencia de tumor tiene dos categorías (Sí, No).

El experimento consiste en "Observar un equino y clasificarlo de acuerdo al color y a la presencia del tumor".

Suponga que se realizó el experimento y las frecuencias observadas que resultan para cada caso son las que se indican en la siguiente tabla

Tabla 3[1]: Distribución de frecuencias según el color y presencia del tumor.

PRESENCIA DE TUMOR COLOR	Si	No	TOTAL
Alazán	220	80	300
Zaino	135	115	250
Tordillo	415	35	450
TOTAL	770	230	1000

Las hipótesis a probar en este caso son

1. H_0: $p_{ij}=p_i \bullet p_j$ $\qquad$ H_1: $p_{ij} \neq p_i \bullet p_j$

que indican

H_0: El color y la presencia de la enfermedad *son independientes*
H_1: El color y la presencia de la enfermedad *no son independientes* (son dependientes).

2. $$\varepsilon = \sum_{i=1}^{k} \frac{(f_{oi} - f_{ei})^2}{f_{ei}}$$

Si vale H_0, $\varepsilon \sim \chi^2_{(c-1)(f-1)}$ central

Si no vale H_0, $\varepsilon \sim \chi^2_{(c-1)(f-1)}$ no central

donde **c** indica la cantidad de columnas (categorías de una de las variables) y **f**: es la cantidad

[1] *Coincide con la Tabla 5 del Capítulo 1.*

de filas (categorías de la otra variable). En este problema los grados de libertad que corresponden son $(2-1)\bullet(2-1)=1$.

3. Si $\alpha=0.01$, la zona de rechazo es $Z=[a,+\infty)$ con $a=\chi^2_{2,0.01}=9.21$ (Tabla E).

4. Para calcular el valor numérico del estadístico, se deben determinar las frecuencias esperadas de cada categoría, utilizando la Definición 1 y bajo el suposición de que la Hipótesis nula es verdadera (o sea que las dos variables son independientes).

Entonces, por ejemplo, para calcular la frecuencia esperada del suceso Alazán (***A***) y Sí (***S***) se procede de la siguiente manera

$$f_{e(AS)} = N \bullet P(AS) = N \bullet P(A) \bullet P(S) = N \bullet \frac{n_A}{N} \bullet \frac{n_S}{N} = \frac{n_A \bullet n_S}{N}$$

donde:

N es el tamaño de la muestra.

n_A y n_S son los totales marginales (número de alazanes y número de animales con tumor, respectivamente).

Esta igualdad es válida por la independencia de los sucesos y por la definición clásica de probabilidad.

Para el Ejemplo 3 las frecuencias esperadas son:

$$f_{e(AS)} = \frac{300 \bullet 770}{1000} = 231 \qquad f_{e(ZS)} = \frac{250 \bullet 770}{1000} = 192.5$$

$$f_{e(AN)} = \frac{300 \bullet 230}{1000} = 69 \qquad f_{e(ZN)} = \frac{250 \bullet 230}{1000} = 57.5$$

$$f_{e(TS)} = \frac{450 \bullet 770}{1000} = 346.5 \qquad f_{e(TN)} = \frac{450 \bullet 230}{1000} = 103.5$$

Reemplazando en la expresión del estadístico las frecuencias observadas y esperadas, se obtiene $\varepsilon_c=135.83$.

5. Dado que el estadístico pertenece a Z, se rechaza la hipótesis nula con probabilidad de cometer error de tipo I de 0.01.

6. Conclusión: *existe relación entre el color de pelaje y presencia de tumor.*

10.3.2. Prueba de Homogeneidad de proporciones

Ejemplo 4: Para comprobar si el uso regular de la aspirina reduce la mortalidad por infarto de miocardio en adultos, se le suministró placebo a 11034 adultos y aspirina a 11037, registrando a lo largo de 5 años si sufrieron infarto de miocardio. En este experimento los adultos no sabían si recibían placebo o aspirina. Los datos obtenidos se presentan en la siguiente tabla.

Tabla 4: Distribución de frecuencias según presencia o no de Infarto y el grupo.

INFARTO DE MIOCARDIO GRUPO	SI	NO	TOTAL
PLACEBO	189	10845	11034
ASPIRINA	104	10933	11037
TOTAL	293	21778	22071

La cantidad de pacientes que recibieron aspirina y placebo está fija de antemano para la experiencia, por lo que se trata de una tabla con los marginales por filas fijos, luego las hipótesis a probar en este caso son

1. H_0: $p_{11} = p_{21}$ H_1: $p_{11} \neq p_{21}$

En palabras estas hipótesis dicen:

H_0: La proporción de los adultos que sufrieron infarto de miocardio es igual para el grupo que recibió placebo y para el que recibió aspirina.

H_1: La proporción de los adultos que sufrieron infarto de miocardio es diferente para el grupo que recibió placebo y para el que recibió aspirina.

2. El estadístico de contraste es idéntico al del caso anterior (expresión y grados de libertad):

$$\varepsilon = \sum_{i=1}^{k} \frac{(f_{oi} - f_{ei})^2}{f_{ei}}$$

Si vale H_0, $\varepsilon \sim \chi^2_{(c-1)(f-1)}$ central

Si no vale H_0, $\varepsilon \sim \chi^2_{(c-1)(f-1)}$ no central

3. Si $\alpha=5\%$, la zona de rechazo es $Z=[a,+\infty)$, donde $a = \chi^2_{1,0.01} = 6.63$ (Tabla E).

4. Las frecuencias teóricas se determinan igual que en el caso anterior:

$$f_{e(PS)} = \frac{293 \bullet 11034}{22071} = 146.48 \qquad f_{e(PN)} = \frac{21778 \bullet 11034}{22071} = 10887.52$$

$$f_{e(AS)} = \frac{293 \bullet 11037}{22071} = 146.52 \qquad f_{e(AN)} = \frac{21778 \bullet 11037}{22071} = 10890.48$$

Entonces $\varepsilon_c=25.014$

5. Como ε_c pertenece a Z se rechaza la hipótesis nula con probabilidad de cometer error de Tipo I de 0.01.

6. Conclusión: *la proporción de adultos que sufrieron infarto de miocardio es diferente para los que recibieron placebo y aspirina.*

La comparación que se realizó entre las proporciones de adultos que *sufrieron infarto*, puede realizarse equivalentemente para aquellos que *no sufrieron infarto*.

10.4 Prueba de Bondad de Ajuste

Para desarrollar esta prueba se utilizará el siguiente:

Ejemplo 5: Se desea saber si la variable X: número de semillas germinadas en paquetes de 4 semillas sigue la distribución binomial. Se sabe que la probabilidad de que una semilla germine es $p=0.70$.

Solución:

Primeramente puede verificarse que el experimento de extraer 4 semillas y ver si germinan cumple los supuestos del modelo Binomial y por tanto se puede pensar que $X_b \sim B(4,0.70)$ siendo el recorrido de la variable $R(X_b)=0,1,2,3,4$. Para comprobar si efectivamente la variable X_b tiene distribución binomial con los parámetros indicados se toma una muestra aleatoria de 150 paquetes de 4 semillas cada uno. Los resultados se indican en la siguiente tabla.

Tabla 5: Distribución de los paquetes de acuerdo al número de semillas que germinan

X_b: **Número de semillas que germinan**	f_o
0	2
1	20
2	42
3	56
4	30
Total	150

Los datos de esta tabla indican, por ejemplo, que de los 150 paquetes hay 2 en los que no germinó ninguna semilla, 20 en los que germinó una semilla, etc.

Se está ahora en condiciones de plantear las siguientes hipótesis

1. H_0: $X_b \sim B(4,0.70)$ H_1: Lo contrario

es decir

H_0: La variable X: Número de semillas germinadas en paquetes de cuatro semillas sigue la distribución Binomial con parámetros $n=4$ y $p=0.70$.

2. El estadístico de contraste es

$$\varepsilon = \sum_{i=1}^{k} \frac{(f_{oi} - f_{ei})^2}{f_{ei}}$$

Si vale H_0, $\varepsilon \sim \chi^2_{k-m-1}$ central

Si no vale H_0, $\varepsilon \sim \chi^2_{k-m-1}$ no central

donde $\mathbf{k}$ indica la cantidad de valores que toma la variable y $\mathbf{m}$ la cantidad de parámetros a estimar.

Para este problema el estadístico ε tiene distribución χ^2_{5-0-1} dado que $\mathbf{m=0}$ porque se conocen los valores de $\mathbf{n}$ y $\mathbf{p}$.

En general el valor de **n** es siempre conocido; en tanto que el parámetro **p** puede ser conocido o no. En caso que sea desconocido puede ser estimado teniendo en cuenta que $E(X_b)=n\bullet p$ y como la media muestral es un buen estimador de la media poblacional (o sea que $E(X_b) \cong \overline{X}$) entonces se puede considerar que $\overline{X} \cong n \bullet p$, de donde $\hat{p} = \frac{\overline{X}}{n}$.

3. Si $\alpha=5\%$, la zona de rechazo es $Z=[a,+\infty)$ con $a = \chi^2_{4,0.05} = 9.49$ (Tabla E), lo que se puede visualizar en el siguiente gráfico.

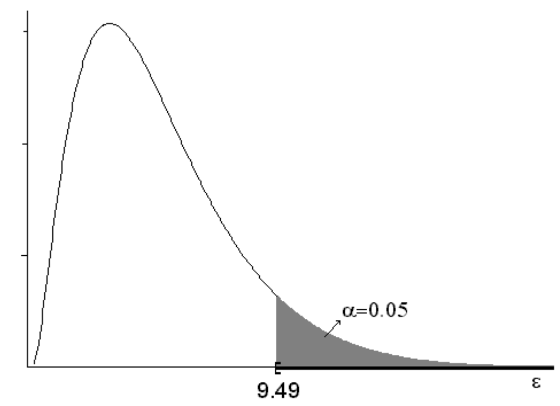


Gráfico 2: Nivel de significación y zona de rechazo para el Ejemplo 5

4. Para calcular el estadístico de contraste se deben determinar las frecuencias esperadas f_e, sobre la base de la Definición 1 y teniendo en cuenta que la variable X_b~$B(4,0.70)$ bajo H_0.

$$f_{e1} = 150 \bullet P(X_b = 0) = 150 \bullet 0.008 = 1.2$$
$$f_{e2} = 150 \bullet P(X_b = 1) = 150 \bullet 0.076 = 11.4$$
$$f_{e3} = 150 \bullet P(X_b = 2) = 150 \bullet 0.265 = 39.75$$
$$f_{e4} = 150 \bullet P(X_b = 3) = 150 \bullet 0.412 = 61.8$$
$$f_{e5} = 150 \bullet P(X_b = 4) = 150 \bullet 0.240 = 36$$

donde los valores de las probabilidades son obtenidos de la Tabla A del Apéndice. A partir de esta información, $\varepsilon_c=8.96$.

5. Como el ε_c no pertenece a Z, la decisión es no rechazar H_0, con probabilidad de cometer Error de tipo **II**.

6. Conclusión: *no hay evidencias para decir que la variable X: número de semillas que germinan entre las cuatro seleccionadas, no sigue la distribución Binomial con parámetros $n=4$ y $p=0.70$.*

En el problema anterior se deseaba ver si la distribución teórica de la variable en estudio era Binomial. También se podría estudiar si una determinada variable aleatoria tiene distribución de Poisson, Normal, etc.; la única diferencia es que la frecuencia esperada depende que distribución planteada en la hipótesis nula (Meyer, P. 1992).

10.5 Conclusiones Finales

1. En problemas donde puede aplicarse la Prueba de Homogeneidad de Proporciones y se tienen sólo dos proporciones a comparar, se puede utilizar también la Prueba de Diferencia de Proporciones (Capítulo 7).

2. Cuando los datos experimentales pueden ser arreglados en tablas de contingencia 2x2 se suelen utilizar algunas medidas de asociación que ayudan a la interpretación de la información experimental. Entre las medidas que se pueden determinar están el *Riesgo Relativo* y los *Odds Ratios*. Para más detalle Agresti, A. (1990) y Ato García, M. y López García, J. (1996).

3. Para el caso de tablas de contingencia 2x2 que tengan todos los totales marginales fijos, se utiliza la *Prueba Exacta de Fisher* (Agresti, A. - 1996).

4. Las frecuencias esperadas no deben ser menores a 1 y no más del 20% menores a cinco. Si esto no se cumple hay que combinar las categorías de las variables para que las frecuencias esperadas tengan la magnitud deseada (Snedecor, G.W. y Cochran, W.G. 1978)

5. Cuando el estadístico tiene distribución Ji-Cuadrado con un solo grado de libertad se considera que la aproximación a esta distribución no es adecuada, por lo que conviene realizar una corrección llamada la *Corrección de Yates*. En base a ésta el estadístico a usar es

$$\varepsilon = \sum_{i=1}^{k} \frac{\left(|fe - fo| - 1/2\right)^2}{fe}$$

6. Las variables cuantitativas pueden categorizarse, no teniendo en cuenta las medidas reales sino sólo las categorías y sus frecuencias.

Ejercicios de Aplicación

1.

En ciertos casos de herencia se ha encontrado que algunas características son heredadas en la proporción 3:1, es decir a la larga tres cuartos de la descendencia tendrán una característica dada y un cuarto no.
En un ensayo se cruzaron plantas con cotiledones amarillos con plantas con cotiledones verdes, observándose que en F2 (segunda generación filial) 6022 plantas tenían cotiledones amarillos y 2001 verdes. Se desea probar (estadísticamente) que el color se hereda en la proporción 3:1 (amarillos, verdes).

a) ¿Cuál sería la prueba adecuada en este caso?

b) Escribir las hipótesis correspondientes.

c) ¿Cuál es el estadístico y cuál es su distribución teórica?

d) Realizar el análisis adecuado y escribir la conclusión.

2.

Se desea saber si un tratamiento es efectivo para curar cierta infección ocular. Para tomar una decisión se eligieron aleatoriamente 200 enfermos, algunos de los cuales recibieron tratamiento y otros no, obteniéndose los siguientes resultados:

	CURADOS	NO-CURADOS
TRATADOS	140	20
NO TRATADOS	10	30

a) ¿Cuáles son las variables que intervienen y a qué tipo corresponden?

b) ¿Cuál es la prueba adecuada para este problema? Indicar las hipótesis correspondientes.

c) Para obtener el valor numérico del estadístico. ¿Qué información es necesaria? Calcularlo.

d) Determinar aproximadamente el valor p del test y obtener conclusiones.

3.

En pacientes con úlcera gástrica se desea establecer si existe relación entre el lugar de la úlcera y el grado de malignidad.

a) ¿Cuál es la prueba adecuada para resolver este problema?

b) Indicar las hipótesis correspondientes.

c) Para tomar una decisión sobre las hipótesis planteadas se observaron 211 pacientes clasificándolos de la siguiente forma:

GRADO DE MALIGNIDAD LUGAR	Benigna	Maligna
Prepilórica	87	34
Cuerpo	52	19
Cardias	11	8

Establecer conclusiones a partir de la muestra.

4.

En estudios genéticos se ha encontrado que el 25% de las moscas de la fruta tienen ojos blancos. En un ensayo se encontró que 1981 moscas de la fruta tenían ojos blancos, mientras que 7712 los tenían rojos ¿Concuerdan estos resultados observados con la proporción teórica?

5.

En un estudio realizado con vacas de distintas razas se deseaba determinar si la fecundidad está asociada con las razas.

a) Decir cuál es la prueba adecuada para resolver este problema.

b) Para comprobar lo anterior se clasificaron los animales de acuerdo a la siguiente tabla:

FECUNDIDAD RAZA	Fecundados	No Fecundados
Charolés	515	1287
Indubrasil	506	665
Nerolé	58	70
Char-Cebú	205	93

Plantear las hipótesis correspondientes y dar la conclusión sabiendo que el valor del estadístico es 204.61 y su correspondiente valor p es 0.000.

6.

En un ensayo se cruzaron arvejas de flores azules (B) y granos de polen alargado (L) con otras de flores rojas (b) y granos de polen redondeados (l). Como estas

características se heredan independientemente, en la segunda generación deberían aparecer las cuatro categorías siguientes BL, Bl, bL, bl en la proporción de Mendel 9:3:3:1. Se observaron 419 plantas encontrándose lo siguiente:

	B	b
L	226	97
l	95	1

a) ¿Qué se desea probar en este caso?

b) El valor del estadístico resultó 32.394 y el valor p=0.0.Tomar la decisión y dar la conclusión en términos del problema.

7.

En una encuesta de salud realizada en la provincia de Tucumán se obtuvo la siguiente distribución del número de hijos varones en familias con 4 hijos.

X_i	0	1	2	3	4
f_i	5	20	44	24	7

Determinar si esta distribución se aparta de la Binomial, suponiendo que ambos sexos son igualmente probables.

8.

Se llevó a cabo un muestreo para estimar el número medio de insectos por parcela cultivada con un cereal. El método de recuento se realizó mediante una red apropiada para el caso.

Los datos siguientes representan la distribución de los resultados muestrales:

X_i	0	1	2	3	4	5	6 o más
f_i	78	167	243	215	135	81	39

a) Sugerir una distribución teórica apropiada de donde provendrían estos datos.

b) De acuerdo a la respuesta dada en a), plantear las hipótesis, el estadístico y sacar las conclusiones sabiendo que el valor p=0.312.

9.

La tabla siguiente muestra la distribución de frecuencias correspondientes a la ganancia de peso en kg. de novillos de una cierta raza.

Intervalos	f_i
[59.5 ; 69.5)	7
[69.5 ; 79.5)	11
[79.5 ; 89.5)	15
[89.5 ; 99.5)	9
[99.5 ; 109.5]	8

¿Puede decir si la distribución de la variable es diferente de una normal?



Tablas Estadísticas

TABLA A :DISTRIBUCIÓN BINOMIAL.

TABLA B :DISTRIBUCIÓN POISSON.

TABLA C :DISTRIBUCIÓN NORMAL ESTÁNDAR.

TABLA D :DISTRIBUCIÓN **t** de STUDENT.

TABLA E :DISTRIBUCIÓN JI-CUADRADO.

TABLA F :DISTRIBUCIÓN F de FISHER.

TABLA A
Distribución Binomial

		p															
n	**k**	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.35**	**0.4**	**0.45**	**0.5**	**0.6**	**0.7**	**0.75**	**0.8**	**0.9**	**0.95**
1	**0**	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000	0.4000	0.3000	0.2500	0.2000	0.1000	0.0500
	1	0.0500	0.1000	0.1500	0.2000	0.2500	0.3000	0.3500	0.4000	0.4500	0.5000	0.6000	0.7000	0.7500	0.8000	0.9000	0.9500
2	**0**	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500	0.1600	0.0900	0.0625	0.0400	0.0100	0.0025
	1	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000	0.4800	0.4200	0.3750	0.3200	0.1800	0.0950
	2	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500	0.3600	0.4900	0.5625	0.6400	0.8100	0.9025
3	**0**	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250	0.0640	0.0270	0.0156	0.0080	0.0010	0.0001
	1	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750	0.2880	0.1890	0.1406	0.0960	0.0270	0.0071
	2	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750	0.4320	0.4410	0.4219	0.3840	0.2430	0.1354
	3	0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250	0.2160	0.3430	0.4219	0.5120	0.7290	0.8574
4	**0**	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625	0.0256	0.0081	0.0039	0.0016	0.0001	0.0000
	1	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500	0.1536	0.0756	0.0469	0.0256	0.0036	0.0005
	2	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750	0.3456	0.2646	0.2109	0.1536	0.0486	0.0135
	3	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500	0.3456	0.4116	0.4219	0.4096	0.2916	0.1715
	4	0.0000	0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625	0.1296	0.2401	0.3164	0.4096	0.6561	0.8145
5	**0**	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313	0.0102	0.0024	0.0010	0.0003	0.0000	0.0000
	1	0.2036	0.3281	0.3915	0.4096	0.3955	0.3602	0.3124	0.2592	0.2059	0.1563	0.0768	0.0284	0.0146	0.0064	0.0005	0.0000
	2	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125	0.2304	0.1323	0.0879	0.0512	0.0081	0.0011
	3	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1811	0.2304	0.2757	0.3125	0.3456	0.3087	0.2637	0.2048	0.0729	0.0214
	4	0.0000	0.0005	0.0022	0.0064	0.0146	0.0284	0.0488	0.0768	0.1128	0.1563	0.2592	0.3602	0.3955	0.4096	0.3281	0.2036
	5	0.0000	0.0000	0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0313	0.0778	0.1681	0.2373	0.3277	0.5905	0.7738
6	**0**	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156	0.0041	0.0007	0.0002	0.0001	0.0000	0.0000
	1	0.2321	0.3543	0.3993	0.3932	0.3560	0.3025	0.2437	0.1866	0.1359	0.0938	0.0369	0.0102	0.0044	0.0015	0.0001	0.0000
	2	0.0305	0.0984	0.1762	0.2458	0.2966	0.3241	0.3280	0.3110	0.2780	0.2344	0.1382	0.0595	0.0330	0.0154	0.0012	0.0001
	3	0.0021	0.0146	0.0415	0.0819	0.1318	0.1852	0.2355	0.2765	0.3032	0.3125	0.2765	0.1852	0.1318	0.0819	0.0146	0.0021
	4	0.0001	0.0012	0.0055	0.0154	0.0330	0.0595	0.0951	0.1382	0.1861	0.2344	0.3110	0.3241	0.2966	0.2458	0.0984	0.0305
	5	0.0000	0.0001	0.0004	0.0015	0.0044	0.0102	0.0205	0.0369	0.0609	0.0938	0.1866	0.3025	0.3560	0.3932	0.3543	0.2321
	6	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0018	0.0041	0.0083	0.0156	0.0467	0.1176	0.1780	0.2621	0.5314	0.7351
7	**0**	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078	0.0016	0.0002	0.0001	0.0000	0.0000	0.0000
	1	0.2573	0.3720	0.3960	0.3670	0.3115	0.2471	0.1848	0.1306	0.0872	0.0547	0.0172	0.0036	0.0013	0.0004	0.0000	0.0000
	2	0.0406	0.1240	0.2097	0.2753	0.3115	0.3177	0.2985	0.2613	0.2140	0.1641	0.0774	0.0250	0.0115	0.0043	0.0002	0.0000
	3	0.0036	0.0230	0.0617	0.1147	0.1730	0.2269	0.2679	0.2903	0.2918	0.2734	0.1935	0.0972	0.0577	0.0287	0.0026	0.0002
	4	0.0002	0.0026	0.0109	0.0287	0.0577	0.0972	0.1442	0.1935	0.2388	0.2734	0.2903	0.2269	0.1730	0.1147	0.0230	0.0036
	5	0.0000	0.0002	0.0012	0.0043	0.0115	0.0250	0.0466	0.0774	0.1172	0.1641	0.2613	0.3177	0.3115	0.2753	0.1240	0.0406
	6	0.0000	0.0000	0.0001	0.0004	0.0013	0.0036	0.0084	0.0172	0.0320	0.0547	0.1306	0.2471	0.3115	0.3670	0.3720	0.2573
	7	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0006	0.0016	0.0037	0.0078	0.0280	0.0824	0.1335	0.2097	0.4783	0.6983
8	**0**	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.2793	0.3826	0.3847	0.3355	0.2670	0.1977	0.1373	0.0896	0.0548	0.0313	0.0079	0.0012	0.0004	0.0001	0.0000	0.0000
	2	0.0515	0.1488	0.2376	0.2936	0.3115	0.2965	0.2587	0.2090	0.1569	0.1094	0.0413	0.0100	0.0038	0.0011	0.0000	0.0000
	3	0.0054	0.0331	0.0839	0.1468	0.2076	0.2541	0.2786	0.2787	0.2568	0.2188	0.1239	0.0467	0.0231	0.0092	0.0004	0.0000
	4	0.0004	0.0046	0.0185	0.0459	0.0865	0.1361	0.1875	0.2322	0.2627	0.2734	0.2322	0.1361	0.0865	0.0459	0.0046	0.0004
	5	0.0000	0.0004	0.0026	0.0092	0.0231	0.0467	0.0808	0.1239	0.1719	0.2188	0.2787	0.2541	0.2076	0.1468	0.0331	0.0054
	6	0.0000	0.0000	0.0002	0.0011	0.0038	0.0100	0.0217	0.0413	0.0703	0.1094	0.2090	0.2965	0.3115	0.2936	0.1488	0.0515
	7	0.0000	0.0000	0.0000	0.0001	0.0004	0.0012	0.0033	0.0079	0.0164	0.0313	0.0896	0.1977	0.2670	0.3355	0.3826	0.2793
	8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0017	0.0039	0.0168	0.0576	0.1001	0.1678	0.4305	0.6634
9	**0**	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.2985	0.3874	0.3679	0.3020	0.2253	0.1556	0.1004	0.0605	0.0339	0.0176	0.0035	0.0004	0.0001	0.0000	0.0000	0.0000
	2	0.0629	0.1722	0.2597	0.3020	0.3003	0.2668	0.2162	0.1612	0.1110	0.0703	0.0212	0.0039	0.0012	0.0003	0.0000	0.0000
	3	0.0077	0.0446	0.1069	0.1762	0.2336	0.2668	0.2716	0.2508	0.2119	0.1641	0.0743	0.0210	0.0087	0.0028	0.0001	0.0000
	4	0.0006	0.0074	0.0283	0.0661	0.1168	0.1715	0.2194	0.2508	0.2600	0.2461	0.1672	0.0735	0.0389	0.0165	0.0008	0.0000
	5	0.0000	0.0008	0.0050	0.0165	0.0389	0.0735	0.1181	0.1672	0.2128	0.2461	0.2508	0.1715	0.1168	0.0661	0.0074	0.0006
	6	0.0000	0.0001	0.0006	0.0028	0.0087	0.0210	0.0424	0.0743	0.1160	0.1641	0.2508	0.2668	0.2336	0.1762	0.0446	0.0077
	7	0.0000	0.0000	0.0000	0.0003	0.0012	0.0039	0.0098	0.0212	0.0407	0.0703	0.1612	0.2668	0.3003	0.3020	0.1722	0.0629
	8	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0013	0.0035	0.0083	0.0176	0.0605	0.1556	0.2253	0.3020	0.3874	0.2985
	9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008	0.0020	0.0101	0.0404	0.0751	0.1342	0.3874	0.6302
10	**0**	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3151	0.3874	0.3474	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098	0.0016	0.0001	0.0000	0.0000	0.0000	0.0000
	2	0.0746	0.1937	0.2759	0.3020	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439	0.0106	0.0014	0.0004	0.0001	0.0000	0.0000
	3	0.0105	0.0574	0.1298	0.2013	0.2503	0.2668	0.2522	0.2150	0.1665	0.1172	0.0425	0.0090	0.0031	0.0008	0.0000	0.0000
	4	0.0010	0.0112	0.0401	0.0881	0.1460	0.2001	0.2377	0.2508	0.2384	0.2051	0.1115	0.0368	0.0162	0.0055	0.0001	0.0000
	5	0.0001	0.0015	0.0085	0.0264	0.0584	0.1029	0.1536	0.2007	0.2340	0.2461	0.2007	0.1029	0.0584	0.0264	0.0015	0.0001
	6	0.0000	0.0001	0.0012	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051	0.2508	0.2001	0.1460	0.0881	0.0112	0.0010
	7	0.0000	0.0000	0.0001	0.0008	0.0031	0.0090	0.0212	0.0425	0.0746	0.1172	0.2150	0.2668	0.2503	0.2013	0.0574	0.0105
	8	0.0000	0.0000	0.0000	0.0001	0.0004	0.0014	0.0043	0.0106	0.0229	0.0439	0.1209	0.2335	0.2816	0.3020	0.1937	0.0746

TABLA A (Continuación)

		p															
n	**k**	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.35**	**0.4**	**0.45**	**0.5**	**0.6**	**0.7**	**0.75**	**0.8**	**0.9**	**0.95**
10	**9**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0016	0.0042	0.0098	0.0403	0.1211	0.1877	0.2684	0.3874	0.3151
	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0.0060	0.0282	0.0563	0.1074	0.3487	0.5987
11	**0**	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3293	0.3835	0.3248	0.2362	0.1549	0.0932	0.0518	0.0266	0.0125	0.0054	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0867	0.2131	0.2866	0.2953	0.2581	0.1998	0.1395	0.0887	0.0513	0.0269	0.0052	0.0005	0.0001	0.0000	0.0000	0.0000
	3	0.0137	0.0710	0.1517	0.2215	0.2581	0.2568	0.2254	0.1774	0.1259	0.0806	0.0234	0.0037	0.0011	0.0002	0.0000	0.0000
	4	0.0014	0.0158	0.0536	0.1107	0.1721	0.2201	0.2428	0.2365	0.2060	0.1611	0.0701	0.0173	0.0064	0.0017	0.0000	0.0000
	5	0.0001	0.0025	0.0132	0.0388	0.0803	0.1321	0.1830	0.2207	0.2360	0.2256	0.1471	0.0566	0.0268	0.0097	0.0003	0.0000
	6	0.0000	0.0003	0.0023	0.0097	0.0268	0.0566	0.0985	0.1471	0.1931	0.2256	0.2207	0.1321	0.0803	0.0388	0.0025	0.0001
	7	0.0000	0.0000	0.0003	0.0017	0.0064	0.0173	0.0379	0.0701	0.1128	0.1611	0.2365	0.2201	0.1721	0.1107	0.0158	0.0014
	8	0.0000	0.0000	0.0000	0.0002	0.0011	0.0037	0.0102	0.0234	0.0462	0.0806	0.1774	0.2568	0.2581	0.2215	0.0710	0.0137
	9	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0018	0.0052	0.0126	0.0269	0.0887	0.1998	0.2581	0.2953	0.2131	0.0867
	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0007	0.0021	0.0054	0.0266	0.0932	0.1549	0.2362	0.3835	0.3293
	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0005	0.0036	0.0198	0.0422	0.0859	0.3138	0.5688
12	**0**	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3413	0.3766	0.3012	0.2062	0.1267	0.0712	0.0368	0.0174	0.0075	0.0029	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0988	0.2301	0.2924	0.2835	0.2323	0.1678	0.1088	0.0639	0.0339	0.0161	0.0025	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.0173	0.0852	0.1720	0.2362	0.2581	0.2397	0.1954	0.1419	0.0923	0.0537	0.0125	0.0015	0.0004	0.0001	0.0000	0.0000
	4	0.0021	0.0213	0.0683	0.1329	0.1936	0.2311	0.2367	0.2128	0.1700	0.1208	0.0420	0.0078	0.0024	0.0005	0.0000	0.0000
	5	0.0002	0.0038	0.0193	0.0532	0.1032	0.1585	0.2039	0.2270	0.2225	0.1934	0.1009	0.0291	0.0115	0.0033	0.0000	0.0000
	6	0.0000	0.0005	0.0040	0.0155	0.0401	0.0792	0.1281	0.1766	0.2124	0.2256	0.1766	0.0792	0.0401	0.0155	0.0005	0.0000
	7	0.0000	0.0000	0.0006	0.0033	0.0115	0.0291	0.0591	0.1009	0.1489	0.1934	0.2270	0.1585	0.1032	0.0532	0.0038	0.0002
	8	0.0000	0.0000	0.0001	0.0005	0.0024	0.0078	0.0199	0.0420	0.0762	0.1208	0.2128	0.2311	0.1936	0.1329	0.0213	0.0021
	9	0.0000	0.0000	0.0000	0.0001	0.0004	0.0015	0.0048	0.0125	0.0277	0.0537	0.1419	0.2397	0.2581	0.2362	0.0852	0.0173
	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0008	0.0025	0.0068	0.0161	0.0639	0.1678	0.2323	0.2835	0.2301	0.0988
	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0.0029	0.0174	0.0712	0.1267	0.2062	0.3766	0.3413
	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0022	0.0138	0.0317	0.0687	0.2824	0.5404
13	**0**	0.5133	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3512	0.3672	0.2774	0.1787	0.1029	0.0540	0.0259	0.0113	0.0045	0.0016	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.1109	0.2448	0.2937	0.2680	0.2059	0.1388	0.0836	0.0453	0.0220	0.0095	0.0012	0.0001	0.0000	0.0000	0.0000	0.0000
	3	0.0214	0.0997	0.1900	0.2457	0.2517	0.2181	0.1651	0.1107	0.0660	0.0349	0.0065	0.0006	0.0001	0.0000	0.0000	0.0000
	4	0.0028	0.0277	0.0838	0.1535	0.2097	0.2337	0.2222	0.1845	0.1350	0.0873	0.0243	0.0034	0.0009	0.0001	0.0000	0.0000
	5	0.0003	0.0055	0.0266	0.0691	0.1258	0.1803	0.2154	0.2214	0.1989	0.1571	0.0656	0.0142	0.0047	0.0011	0.0000	0.0000
	6	0.0000	0.0008	0.0063	0.0230	0.0559	0.1030	0.1546	0.1968	0.2169	0.2095	0.1312	0.0442	0.0186	0.0058	0.0001	0.0000
	7	0.0000	0.0001	0.0011	0.0058	0.0186	0.0442	0.0833	0.1312	0.1775	0.2095	0.1968	0.1030	0.0559	0.0230	0.0008	0.0000
	8	0.0000	0.0000	0.0001	0.0011	0.0047	0.0142	0.0336	0.0656	0.1089	0.1571	0.2214	0.1803	0.1258	0.0691	0.0055	0.0003
	9	0.0000	0.0000	0.0000	0.0001	0.0009	0.0034	0.0101	0.0243	0.0495	0.0873	0.1845	0.2337	0.2097	0.1535	0.0277	0.0028
	10	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0022	0.0065	0.0162	0.0349	0.1107	0.2181	0.2517	0.2457	0.0997	0.0214
	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0012	0.0036	0.0095	0.0453	0.1388	0.2059	0.2680	0.2448	0.1109
	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0016	0.0113	0.0540	0.1029	0.1787	0.3672	0.3512
	13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0013	0.0097	0.0238	0.0550	0.2542	0.5133
14	**0**	0.4877	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3593	0.3559	0.2539	0.1539	0.0832	0.0407	0.0181	0.0073	0.0027	0.0009	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.1229	0.2570	0.2912	0.2501	0.1802	0.1134	0.0634	0.0317	0.0141	0.0056	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0259	0.1142	0.2056	0.2501	0.2402	0.1943	0.1366	0.0845	0.0462	0.0222	0.0033	0.0002	0.0000	0.0000	0.0000	0.0000
	4	0.0037	0.0349	0.0998	0.1720	0.2202	0.2290	0.2022	0.1549	0.1040	0.0611	0.0136	0.0014	0.0003	0.0000	0.0000	0.0000
	5	0.0004	0.0078	0.0352	0.0860	0.1468	0.1963	0.2178	0.2066	0.1701	0.1222	0.0408	0.0066	0.0018	0.0003	0.0000	0.0000
	6	0.0000	0.0013	0.0093	0.0322	0.0734	0.1262	0.1759	0.2066	0.2088	0.1833	0.0918	0.0232	0.0082	0.0020	0.0000	0.0000
	7	0.0000	0.0002	0.0019	0.0092	0.0280	0.0618	0.1082	0.1574	0.1952	0.2095	0.1574	0.0618	0.0280	0.0092	0.0002	0.0000
	8	0.0000	0.0000	0.0003	0.0020	0.0082	0.0232	0.0510	0.0918	0.1398	0.1833	0.2066	0.1262	0.0734	0.0322	0.0013	0.0000
	9	0.0000	0.0000	0.0000	0.0003	0.0018	0.0066	0.0183	0.0408	0.0762	0.1222	0.2066	0.1963	0.1468	0.0860	0.0078	0.0004
	10	0.0000	0.0000	0.0000	0.0000	0.0003	0.0014	0.0049	0.0136	0.0312	0.0611	0.1549	0.2290	0.2202	0.1720	0.0349	0.0037
	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0010	0.0033	0.0093	0.0222	0.0845	0.1943	0.2402	0.2501	0.1142	0.0259
	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0019	0.0056	0.0317	0.1134	0.1802	0.2501	0.2570	0.1229
	13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0009	0.0073	0.0407	0.0832	0.1539	0.3559	0.3593
	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0008	0.0068	0.0178	0.0440	0.2288	0.4877

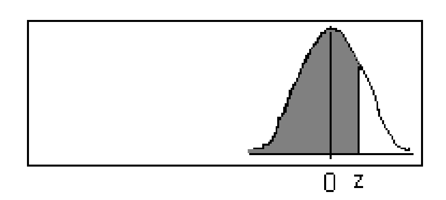
TABLA B
Distribución de Poisson

x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
λ																									
0.1	0.9048	0.0905	0.0045	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	0.8187	0.1637	0.0164	0.0011	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.3	0.7408	0.2222	0.0333	0.0033	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.4	0.6703	0.2681	0.0536	0.0072	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.5	0.6065	0.3033	0.0758	0.0126	0.0016	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.6	0.5488	0.3293	0.0988	0.0198	0.0030	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.7	0.4966	0.3476	0.1217	0.0284	0.0050	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.8	0.4493	0.3595	0.1438	0.0383	0.0077	0.0012	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.9	0.4066	0.3659	0.1647	0.0494	0.0111	0.0020	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.3679	0.3679	0.1839	0.0613	0.0153	0.0031	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.1	0.3329	0.3662	0.2014	0.0738	0.0203	0.0045	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.2	0.3012	0.3614	0.2169	0.0867	0.0260	0.0062	0.0012	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.3	0.2725	0.3543	0.2303	0.0998	0.0324	0.0084	0.0018	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.4	0.2466	0.3452	0.2417	0.1128	0.0395	0.0111	0.0026	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.5	0.2231	0.3347	0.2510	0.1255	0.0471	0.0141	0.0035	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.6	0.2019	0.3230	0.2584	0.1378	0.0551	0.0176	0.0047	0.0011	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.7	0.1827	0.3106	0.2640	0.1496	0.0636	0.0216	0.0061	0.0015	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.8	0.1653	0.2975	0.2678	0.1607	0.0723	0.0260	0.0078	0.0020	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.9	0.1496	0.2842	0.2700	0.1710	0.0812	0.0309	0.0098	0.0027	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.0120	0.0034	0.0009	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2.2	0.1108	0.2438	0.2681	0.1966	0.1082	0.0476	0.0174	0.0055	0.0015	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2.4	0.0907	0.2177	0.2613	0.2090	0.1254	0.0602	0.0241	0.0083	0.0025	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2.6	0.0743	0.1931	0.2510	0.2176	0.1414	0.0735	0.0319	0.0118	0.0038	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2.8	0.0608	0.1703	0.2384	0.2225	0.1557	0.0872	0.0407	0.0163	0.0057	0.0018	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0498	0.1494	0.2240	0.2240	0.1680	0.1008	0.0504	0.0216	0.0081	0.0027	0.0008	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3.2	0.0408	0.1304	0.2087	0.2226	0.1781	0.1140	0.0608	0.0278	0.0111	0.0040	0.0013	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3.4	0.0334	0.1135	0.1929	0.2186	0.1858	0.1264	0.0716	0.0348	0.0148	0.0056	0.0019	0.0006	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3.6	0.0273	0.0984	0.1771	0.2125	0.1912	0.1377	0.0826	0.0425	0.0191	0.0076	0.0028	0.0009	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3.8	0.0224	0.0850	0.1615	0.2046	0.1944	0.1477	0.0936	0.0508	0.0241	0.0102	0.0039	0.0013	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0183	0.0733	0.1465	0.1954	0.1954	0.1563	0.1042	0.0595	0.0298	0.0132	0.0053	0.0019	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0067	0.0337	0.0842	0.1404	0.1755	0.1755	0.1462	0.1044	0.0653	0.0363	0.0181	0.0082	0.0034	0.0013	0.0005	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0025	0.0149	0.0446	0.0892	0.1339	0.1606	0.1606	0.1377	0.1033	0.0688	0.0413	0.0225	0.0113	0.0052	0.0022	0.0009	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0009	0.0064	0.0223	0.0521	0.0912	0.1277	0.1490	0.1490	0.1304	0.1014	0.0710	0.0452	0.0263	0.0142	0.0071	0.0033	0.0014	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0003	0.0027	0.0107	0.0286	0.0573	0.0916	0.1221	0.1396	0.1396	0.1241	0.0993	0.0722	0.0481	0.0296	0.0169	0.0090	0.0045	0.0021	0.0009	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000
9	0.0001	0.0011	0.0050	0.0150	0.0337	0.0607	0.0911	0.1171	0.1318	0.1318	0.1186	0.0970	0.0728	0.0504	0.0324	0.0194	0.0109	0.0058	0.0029	0.0014	0.0006	0.0003	0.0001	0.0000	0.0000
10	0.0000	0.0005	0.0023	0.0076	0.0189	0.0378	0.0631	0.0901	0.1126	0.1251	0.1251	0.1137	0.0948	0.0729	0.0521	0.0347	0.0217	0.0128	0.0071	0.0037	0.0019	0.0009	0.0004	0.0002	0.0001
11	0.0000	0.0002	0.0010	0.0037	0.0102	0.0224	0.0411	0.0646	0.0888	0.1085	0.1194	0.1194	0.1094	0.0926	0.0728	0.0534	0.0367	0.0237	0.0145	0.0084	0.0046	0.0024	0.0012	0.0006	0.0003
12	0.0000	0.0001	0.0004	0.0018	0.0053	0.0127	0.0255	0.0437	0.0655	0.0874	0.1048	0.1144	0.1144	0.1056	0.0905	0.0724	0.0543	0.0383	0.0255	0.0161	0.0097	0.0055	0.0030	0.0016	0.0008
13	0.0000	0.0000	0.0002	0.0008	0.0027	0.0070	0.0152	0.0281	0.0457	0.0661	0.0859	0.1015	0.1099	0.1099	0.1021	0.0885	0.0719	0.0550	0.0397	0.0272	0.0177	0.0109	0.0065	0.0037	0.0020
14	0.0000	0.0000	0.0001	0.0004	0.0013	0.0037	0.0087	0.0174	0.0304	0.0473	0.0663	0.0844	0.0984	0.1060	0.1060	0.0989	0.0866	0.0713	0.0554	0.0409	0.0286	0.0191	0.0121	0.0074	0.0043

TABLA C

Distribución Normal Estándar (N(0,1))

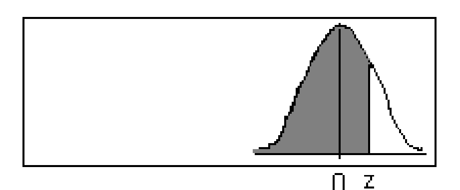
Esta tabla presenta la probabilidad acumulada $P(Z \leq z)$, para $z \geq 0$



z	P(Z<z)	z	P(Z<z)	z	P(Z<z)	z	P(Z<z)	z	P(Z<z)	z	P(Z<z)	z	P(Z<z)	z	P(Z<z)
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772	2.50	0.9938	3.00	0.9987	3.50	0.9998
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.01	0.9778	2.51	0.9940	3.01	0.9987	3.51	0.9998
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.02	0.9783	2.52	0.9941	3.02	0.9987	3.52	0.9998
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.03	0.9788	2.53	0.9943	3.03	0.9988	3.53	0.9998
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.04	0.9793	2.54	0.9945	3.04	0.9988	3.54	0.9998
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.05	0.9798	2.55	0.9946	3.05	0.9989	3.55	0.9998
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.06	0.9803	2.56	0.9948	3.06	0.9989	3.56	0.9998
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.07	0.9808	2.57	0.9949	3.07	0.9989	3.57	0.9998
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.08	0.9812	2.58	0.9951	3.08	0.9990	3.58	0.9998
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.09	0.9817	2.59	0.9952	3.09	0.9990	3.59	0.9998
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	2.10	0.9821	2.60	0.9953	3.10	0.9990	3.60	0.9998
0.11	0.5438	0.61	0.7291	1.11	0.8665	1.61	0.9463	2.11	0.9826	2.61	0.9955	3.11	0.9991	3.61	0.9998
0.12	0.5478	0.62	0.7324	1.12	0.8686	1.62	0.9474	2.12	0.9830	2.62	0.9956	3.12	0.9991	3.62	0.9999
0.13	0.5517	0.63	0.7357	1.13	0.8708	1.63	0.9484	2.13	0.9834	2.63	0.9957	3.13	0.9991	3.63	0.9999
0.14	0.5557	0.64	0.7389	1.14	0.8729	1.64	0.9495	2.14	0.9838	2.64	0.9959	3.14	0.9992	3.64	0.9999
0.15	0.5596	0.65	0.7422	1.15	0.8749	1.65	0.9505	2.15	0.9842	2.65	0.9960	3.15	0.9992	3.65	0.9999
0.16	0.5636	0.66	0.7454	1.16	0.8770	1.66	0.9515	2.16	0.9846	2.66	0.9961	3.16	0.9992	3.66	0.9999
0.17	0.5675	0.67	0.7486	1.17	0.8790	1.67	0.9525	2.17	0.9850	2.67	0.9962	3.17	0.9992	3.67	0.9999
0.18	0.5714	0.68	0.7517	1.18	0.8810	1.68	0.9535	2.18	0.9854	2.68	0.9963	3.18	0.9993	3.68	0.9999
0.19	0.5753	0.69	0.7549	1.19	0.8830	1.69	0.9545	2.19	0.9857	2.69	0.9964	3.19	0.9993	3.69	0.9999
0.20	0.5793	0.70	0.7580	1.20	0.8849	1.70	0.9554	2.20	0.9861	2.70	0.9965	3.20	0.9993	3.70	0.9999
0.21	0.5832	0.71	0.7611	1.21	0.8869	1.71	0.9564	2.21	0.9864	2.71	0.9966	3.21	0.9993	3.71	0.9999
0.22	0.5871	0.72	0.7642	1.22	0.8888	1.72	0.9573	2.22	0.9868	2.72	0.9967	3.22	0.9994	3.72	0.9999
0.23	0.5910	0.73	0.7673	1.23	0.8907	1.73	0.9582	2.23	0.9871	2.73	0.9968	3.23	0.9994	3.73	0.9999
0.24	0.5948	0.74	0.7704	1.24	0.8925	1.74	0.9591	2.24	0.9875	2.74	0.9969	3.24	0.9994	3.74	0.9999
0.25	0.5987	0.75	0.7734	1.25	0.8944	1.75	0.9599	2.25	0.9878	2.75	0.9970	3.25	0.9994	3.75	0.9999
0.26	0.6026	0.76	0.7764	1.26	0.8962	1.76	0.9608	2.26	0.9881	2.76	0.9971	3.26	0.9994	3.76	0.9999
0.27	0.6064	0.77	0.7794	1.27	0.8980	1.77	0.9616	2.27	0.9884	2.77	0.9972	3.27	0.9995	3.77	0.9999
0.28	0.6103	0.78	0.7823	1.28	0.8997	1.78	0.9625	2.28	0.9887	2.78	0.9973	3.28	0.9995	3.78	0.9999
0.29	0.6141	0.79	0.7852	1.29	0.9015	1.79	0.9633	2.29	0.9890	2.79	0.9974	3.29	0.9995	3.79	0.9999
0.30	0.6179	0.80	0.7881	1.30	0.9032	1.80	0.9641	2.30	0.9893	2.80	0.9974	3.30	0.9995	3.80	0.9999
0.31	0.6217	0.81	0.7910	1.31	0.9049	1.81	0.9649	2.31	0.9896	2.81	0.9975	3.31	0.9995	3.81	0.9999
0.32	0.6255	0.82	0.7939	1.32	0.9066	1.82	0.9656	2.32	0.9898	2.82	0.9976	3.32	0.9995	3.82	0.9999
0.33	0.6293	0.83	0.7967	1.33	0.9082	1.83	0.9664	2.33	0.9901	2.83	0.9977	3.33	0.9996	3.83	0.9999
0.34	0.6331	0.84	0.7995	1.34	0.9099	1.84	0.9671	2.34	0.9904	2.84	0.9977	3.34	0.9996	3.84	0.9999
0.35	0.6368	0.85	0.8023	1.35	0.9115	1.85	0.9678	2.35	0.9906	2.85	0.9978	3.35	0.9996	3.85	0.9999
0.36	0.6406	0.86	0.8051	1.36	0.9131	1.86	0.9686	2.36	0.9909	2.86	0.9979	3.36	0.9996	3.86	0.9999
0.37	0.6443	0.87	0.8078	1.37	0.9147	1.87	0.9693	2.37	0.9911	2.87	0.9979	3.37	0.9996	3.87	0.9999
0.38	0.6480	0.88	0.8106	1.38	0.9162	1.88	0.9699	2.38	0.9913	2.88	0.9980	3.38	0.9996	3.88	0.9999
0.39	0.6517	0.89	0.8133	1.39	0.9177	1.89	0.9706	2.39	0.9916	2.89	0.9981	3.39	0.9997	3.89	0.9999
0.40	0.6554	0.90	0.8159	1.40	0.9192	1.90	0.9713	2.40	0.9918	2.90	0.9981	3.40	0.9997	3.90	1.0000
0.41	0.6591	0.91	0.8186	1.41	0.9207	1.91	0.9719	2.41	0.9920	2.91	0.9982	3.41	0.9997	3.91	1.0000
0.42	0.6628	0.92	0.8212	1.42	0.9222	1.92	0.9726	2.42	0.9922	2.92	0.9982	3.42	0.9997	3.92	1.0000
0.43	0.6664	0.93	0.8238	1.43	0.9236	1.93	0.9732	2.43	0.9925	2.93	0.9983	3.43	0.9997	3.93	1.0000
0.44	0.6700	0.94	0.8264	1.44	0.9251	1.94	0.9738	2.44	0.9927	2.94	0.9984	3.44	0.9997	3.94	1.0000
0.45	0.6736	0.95	0.8289	1.45	0.9265	1.95	0.9744	2.45	0.9929	2.95	0.9984	3.45	0.9997	3.95	1.0000
0.46	0.6772	0.96	0.8315	1.46	0.9279	1.96	0.9750	2.46	0.9931	2.96	0.9985	3.46	0.9997	3.96	1.0000
0.47	0.6808	0.97	0.8340	1.47	0.9292	1.97	0.9756	2.47	0.9932	2.97	0.9985	3.47	0.9997	3.97	1.0000
0.48	0.6844	0.98	0.8365	1.48	0.9306	1.98	0.9761	2.48	0.9934	2.98	0.9986	3.48	0.9997	3.98	1.0000
0.49	0.6879	0.99	0.8389	1.49	0.9319	1.99	0.9767	2.49	0.9936	2.99	0.9986	3.49	0.9998	3.99	1.0000

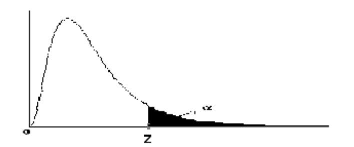
TABLA D
Distribución t de Student

En el margen superior se leen los percentiles y en margen izquierdo los grados de libertad.
Esta tabla da los valores de z > 0 para $P(t_n \leq z)=1-\alpha$

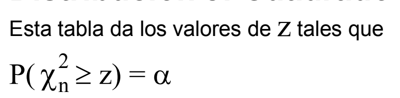


$1-\alpha$	0.995	0.99	0.975	0.95	0.90	0.80	0.75
n							
1	63.66	31.82	12.71	6.31	3.08	1.38	1.00
2	9.92	6.96	4.30	2.92	1.89	1.06	0.82
3	5.84	4.54	3.18	2.35	1.64	0.98	0.76
4	4.60	3.75	2.78	2.13	1.53	0.94	0.74
5	4.03	3.36	2.57	2.02	1.48	0.92	0.73
6	3.71	3.14	2.45	1.94	1.44	0.91	0.72
7	3.50	3.00	2.36	1.89	1.41	0.90	0.71
8	3.36	2.90	2.31	1.86	1.40	0.89	0.71
9	3.25	2.82	2.26	1.83	1.38	0.88	0.70
10	3.17	2.76	2.23	1.81	1.37	0.88	0.70
11	3.11	2.72	2.20	1.80	1.36	0.88	0.70
12	3.05	2.68	2.18	1.78	1.36	0.87	0.70
13	3.01	2.65	2.16	1.77	1.35	0.87	0.69
14	2.98	2.62	2.14	1.76	1.35	0.87	0.69
15	2.95	2.60	2.13	1.75	1.34	0.87	0.69
16	2.92	2.58	2.12	1.75	1.34	0.86	0.69
17	2.90	2.57	2.11	1.74	1.33	0.86	0.69
18	2.88	2.55	2.10	1.73	1.33	0.86	0.69
19	2.86	2.54	2.09	1.73	1.33	0.86	0.69
20	2.85	2.53	2.09	1.72	1.33	0.86	0.69
21	2.83	2.52	2.08	1.72	1.32	0.86	0.69
22	2.82	2.51	2.07	1.72	1.32	0.86	0.69
23	2.81	2.50	2.07	1.71	1.32	0.86	0.69
24	2.80	2.49	2.06	1.71	1.32	0.86	0.68
25	2.79	2.49	2.06	1.71	1.32	0.86	0.68
26	2.78	2.48	2.06	1.71	1.31	0.86	0.68
27	2.77	2.47	2.05	1.70	1.31	0.86	0.68
28	2.76	2.47	2.05	1.70	1.31	0.85	0.68
29	2.76	2.46	2.05	1.70	1.31	0.85	0.68
30	2.75	2.46	2.04	1.70	1.31	0.85	0.68
31	2.74	2.45	2.04	1.70	1.31	0.85	0.68
32	2.74	2.45	2.04	1.69	1.31	0.85	0.68
33	2.73	2.44	2.03	1.69	1.31	0.85	0.68
34	2.73	2.44	2.03	1.69	1.31	0.85	0.68
35	2.72	2.44	2.03	1.69	1.31	0.85	0.68
36	2.72	2.43	2.03	1.69	1.31	0.85	0.68
37	2.72	2.43	2.03	1.69	1.30	0.85	0.68
38	2.71	2.43	2.02	1.69	1.30	0.85	0.68
39	2.71	2.43	2.02	1.68	1.30	0.85	0.68
40	2.70	2.42	2.02	1.68	1.30	0.85	0.68
45	2.69	2.41	2.01	1.68	1.30	0.85	0.68
50	2.68	2.40	2.01	1.68	1.30	0.85	0.68
55	2.67	2.40	2.00	1.67	1.30	0.85	0.68
60	2.66	2.39	2.00	1.67	1.30	0.85	0.68
65	2.65	2.39	2.00	1.67	1.29	0.85	0.68
70	2.65	2.38	1.99	1.67	1.29	0.85	0.68
75	2.64	2.38	1.99	1.67	1.29	0.85	0.68
80	2.64	2.37	1.99	1.66	1.29	0.85	0.68
90	2.63	2.37	1.99	1.66	1.29	0.85	0.68
100	2.63	2.36	1.98	1.66	1.29	0.85	0.68
120	2.62	2.36	1.98	1.66	1.29	0.84	0.68

TABLA E
Distribución Ji-Cuadrado



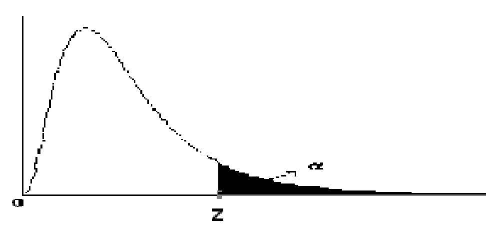
Esta tabla da los valores de Z tales que



$$P(\chi^2_n \geq z) = \alpha$$

α	0.005	0.01	0.025	0.05	0.10	0.20	0.4	0.50	0.75	0.90	0.95	0.995
n												
1	7.88	6.63	5.02	3.84	2.71	1.64	0.71	0.45	0.10	0.02	0.00	0.00
2	10.60	9.21	7.38	5.99	4.61	3.22	1.83	1.39	0.58	0.21	0.10	0.01
3	12.84	11.34	9.35	7.81	6.25	4.64	2.95	2.37	1.21	0.58	0.35	0.07
4	14.86	13.28	11.14	9.49	7.78	5.99	4.04	3.36	1.92	1.06	0.71	0.21
5	16.75	15.09	12.83	11.07	9.24	7.29	5.13	4.35	2.67	1.61	1.15	0.41
6	18.55	16.81	14.45	12.59	10.64	8.56	6.21	5.35	3.45	2.20	1.64	0.68
7	20.28	18.48	16.01	14.07	12.02	9.80	7.28	6.35	4.25	2.83	2.17	0.99
8	21.95	20.09	17.53	15.51	13.36	11.03	8.35	7.34	5.07	3.49	2.73	1.34
9	23.59	21.67	19.02	16.92	14.68	12.24	9.41	8.34	5.90	4.17	3.33	1.73
10	25.19	23.21	20.48	18.31	15.99	13.44	10.47	9.34	6.74	4.87	3.94	2.16
11	26.76	24.73	21.92	19.68	17.28	14.63	11.53	10.34	7.58	5.58	4.57	2.60
12	28.30	26.22	23.34	21.03	18.55	15.81	12.58	11.34	8.44	6.30	5.23	3.07
13	29.82	27.69	24.74	22.36	19.81	16.98	13.64	12.34	9.30	7.04	5.89	3.57
14	31.32	29.14	26.12	23.68	21.06	18.15	14.69	13.34	10.17	7.79	6.57	4.07
15	32.80	30.58	27.49	25.00	22.31	19.31	15.73	14.34	11.04	8.55	7.26	4.60
16	34.27	32.00	28.85	26.30	23.54	20.47	16.78	15.34	11.91	9.31	7.96	5.14
17	35.72	33.41	30.19	27.59	24.77	21.61	17.82	16.34	12.79	10.09	8.67	5.70
18	37.16	34.81	31.53	28.87	25.99	22.76	18.87	17.34	13.68	10.86	9.39	6.26
19	38.58	36.19	32.85	30.14	27.20	23.90	19.91	18.34	14.56	11.65	10.12	6.84
20	40.00	37.57	34.17	31.41	28.41	25.04	20.95	19.34	15.45	12.44	10.85	7.43
21	41.40	38.93	35.48	32.67	29.62	26.17	21.99	20.34	16.34	13.24	11.59	8.03
22	42.80	40.29	36.78	33.92	30.81	27.30	23.03	21.34	17.24	14.04	12.34	8.64
23	44.18	41.64	38.08	35.17	32.01	28.43	24.07	22.34	18.14	14.85	13.09	9.26
24	45.56	42.98	39.36	36.42	33.20	29.55	25.11	23.34	19.04	15.66	13.85	9.89
25	46.93	44.31	40.65	37.65	34.38	30.68	26.14	24.34	19.94	16.47	14.61	10.52
26	48.29	45.64	41.92	38.89	35.56	31.79	27.18	25.34	20.84	17.29	15.38	11.16
27	49.65	46.96	43.19	40.11	36.74	32.91	28.21	26.34	21.75	18.11	16.15	11.81
28	50.99	48.28	44.46	41.34	37.92	34.03	29.25	27.34	22.66	18.94	16.93	12.46
29	52.34	49.59	45.72	42.56	39.09	35.14	30.28	28.34	23.57	19.77	17.71	13.12
30	53.67	50.89	46.98	43.77	40.26	36.25	31.32	29.34	24.48	20.60	18.49	13.79
31	55.00	52.19	48.23	44.99	41.42	37.36	32.35	30.34	25.39	21.43	19.28	14.46
32	56.33	53.49	49.48	46.19	42.58	38.47	33.38	31.34	26.30	22.27	20.07	15.13
33	57.65	54.78	50.73	47.40	43.75	39.57	34.41	32.34	27.22	23.11	20.87	15.82
34	58.96	56.06	51.97	48.60	44.90	40.68	35.44	33.34	28.14	23.95	21.66	16.50
35	60.27	57.34	53.20	49.80	46.06	41.78	36.47	34.34	29.05	24.80	22.47	17.19
36	61.58	58.62	54.44	51.00	47.21	42.88	37.50	35.34	29.97	25.64	23.27	17.89
37	62.88	59.89	55.67	52.19	48.36	43.98	38.53	36.34	30.89	26.49	24.07	18.59
38	64.18	61.16	56.90	53.38	49.51	45.08	39.56	37.34	31.81	27.34	24.88	19.29
39	65.48	62.43	58.12	54.57	50.66	46.17	40.59	38.34	32.74	28.20	25.70	20.00
40	66.77	63.69	59.34	55.76	51.81	47.27	41.62	39.34	33.66	29.05	26.51	20.71
45	73.17	69.96	65.41	61.66	57.51	52.73	46.76	44.34	38.29	33.35	30.61	24.31
50	79.49	76.15	71.42	67.50	63.17	58.16	51.89	49.33	42.94	37.69	34.76	27.99
55	85.75	82.29	77.38	73.31	68.80	63.58	57.02	54.33	47.61	42.06	38.96	31.73
60	91.95	88.38	83.30	79.08	74.40	68.97	62.13	59.33	52.29	46.46	43.19	35.53
65	98.10	94.42	89.18	84.82	79.97	74.35	67.25	64.33	56.99	50.88	47.45	39.38
70	104.21	100.43	95.02	90.53	85.53	79.71	72.36	69.33	61.70	55.33	51.74	43.28
75	110.29	106.39	100.84	96.22	91.06	85.07	77.46	74.33	66.42	59.79	56.05	47.21
80	116.32	112.33	106.63	101.88	96.58	90.41	82.57	79.33	71.14	64.28	60.39	51.17
90	128.30	124.12	118.14	113.15	107.57	101.05	92.76	89.33	80.62	73.29	69.13	59.20
100	140.17	135.81	129.56	124.34	118.50	111.67	102.95	99.33	90.13	82.36	77.93	67.33
120	163.65	158.95	152.21	146.57	140.23	132.81	123.29	119.33	109.22	100.62	95.70	83.85

TABLA F
Distribución F de Fisher



En las columnas grados de libertad del numerador, en las filas grados de libertad del denominador.
Esta tabla da valores de z tales que $P(F_{n_1,n_2} \geq z) = \alpha$

		n 1																							
n 2	**α**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**30**	**40**	**60**	**120**
1	0.005	16212.46	19997.36	21614.13	22500.75	23055.82	23439.53	23715.20	23923.81	24091.45	24221.84	24333.60	24426.73	24504.96	24572.01	24631.62	24683.77	24728.48	24765.73	24802.98	24836.51	25041.40	25145.71	25253.74	25358.05
1	0.01	4052.18	4999.34	5403.53	5624.26	5763.96	5858.95	5928.33	5980.95	6022.40	6055.93	6083.40	6106.68	6125.77	6143.00	6156.97	6170.01	6181.19	6191.43	6200.75	6208.66	6260.35	6286.43	6312.97	6339.51
1	0.025	647.79	799.48	864.15	899.60	921.83	937.11	948.20	956.64	963.28	968.63	973.03	976.72	979.84	982.55	984.87	986.91	988.72	990.35	991.80	993.08	1001.40	1005.60	1009.79	1014.04
1	0.05	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.90	244.69	245.36	245.95	246.47	246.92	247.32	247.69	248.02	250.10	251.14	252.20	253.25
1	0.975	0.00	0.03	0.06	0.08	0.10	0.11	0.12	0.13	0.14	0.14	0.15	0.15	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.18	0.18	0.19	0.19
1	0.95	0.01	0.05	0.10	0.13	0.15	0.17	0.18	0.19	0.20	0.20	0.21	0.21	0.21	0.22	0.22	0.22	0.22	0.23	0.23	0.23	0.24	0.24	0.25	0.26
1	0.995	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.07	0.08	0.08	0.09	0.09	0.09	0.09	0.09	0.10	0.10	0.10	0.10	0.11	0.11	0.12	0.12
2	0.005	198.50	199.01	199.16	199.24	199.30	199.33	199.36	199.38	199.39	199.39	199.42	199.42	199.42	199.42	199.43	199.45	199.45	199.45	199.45	199.45	199.48	199.48	199.48	199.49
2	0.01	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.41	99.42	99.42	99.43	99.43	99.44	99.44	99.44	99.45	99.45	99.47	99.48	99.48	99.49
2	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.41	39.42	39.43	39.43	39.44	39.44	39.44	39.45	39.45	39.46	39.47	39.48	39.49
2	0.05	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42	19.43	19.43	19.44	19.44	19.44	19.45	19.46	19.47	19.48	19.49
2	0.975	0.00	0.03	0.06	0.09	0.12	0.14	0.15	0.17	0.17	0.18	0.19	0.20	0.20	0.21	0.21	0.21	0.22	0.22	0.22	0.22	0.24	0.25	0.25	0.26
2	0.95	0.01	0.05	0.10	0.14	0.17	0.19	0.21	0.22	0.23	0.24	0.25	0.26	0.26	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.30	0.31	0.32	0.33
2	0.995	0.00	0.01	0.02	0.04	0.05	0.07	0.08	0.09	0.10	0.11	0.11	0.12	0.12	0.13	0.13	0.13	0.14	0.14	0.14	0.14	0.16	0.16	0.17	0.18
3	0.005	55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.88	43.68	43.52	43.39	43.27	43.17	43.08	43.01	42.94	42.88	42.83	42.78	42.47	42.31	42.15	41.99
3	0.01	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05	26.98	26.92	26.87	26.83	26.79	26.75	26.72	26.69	26.50	26.41	26.32	26.22
3	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.37	14.34	14.30	14.28	14.25	14.23	14.21	14.20	14.18	14.17	14.08	14.04	13.99	13.95
3	0.05	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67	8.67	8.66	8.62	8.59	8.57	8.55
3	0.975	0.00	0.03	0.06	0.10	0.13	0.15	0.17	0.18	0.20	0.21	0.22	0.22	0.23	0.24	0.24	0.25	0.25	0.25	0.26	0.26	0.28	0.29	0.30	0.31
3	0.95	0.00	0.05	0.11	0.15	0.18	0.21	0.23	0.25	0.26	0.27	0.28	0.29	0.29	0.30	0.30	0.31	0.31	0.32	0.32	0.32	0.34	0.35	0.36	0.37
3	0.995	0.00	0.01	0.02	0.04	0.06	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.14	0.15	0.15	0.16	0.16	0.17	0.17	0.17	0.19	0.20	0.21	0.22
4	0.005	31.33	26.28	24.26	23.15	22.46	21.98	21.62	21.35	21.14	20.97	20.82	20.70	20.60	20.51	20.44	20.37	20.31	20.26	20.21	20.17	19.89	19.75	19.61	19.47

TABLA F (Continuación)

		n_1																							
n_2	α	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	30	40	60	120
4	0.01	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.45	14.37	14.31	14.25	14.20	14.15	14.11	14.08	14.05	14.02	13.84	13.75	13.65	13.56
4	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.79	8.75	8.72	8.68	8.66	8.63	8.61	8.59	8.58	8.56	8.46	8.41	8.36	8.31
4	0.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82	5.81	5.80	5.75	5.72	5.69	5.66
4	0.975	0.00	0.03	0.07	0.10	0.14	0.16	0.18	0.20	0.21	0.22	0.23	0.24	0.25	0.26	0.26	0.27	0.27	0.28	0.28	0.28	0.31	0.32	0.33	0.35
4	0.95	0.00	0.05	0.11	0.16	0.19	0.22	0.24	0.26	0.28	0.29	0.30	0.31	0.31	0.32	0.33	0.33	0.34	0.34	0.35	0.35	0.37	0.38	0.40	0.41
4	0.995	0.00	0.01	0.02	0.04	0.06	0.08	0.10	0.11	0.13	0.14	0.15	0.15	0.16	0.17	0.17	0.18	0.18	0.19	0.19	0.19	0.22	0.23	0.24	0.26
5	0.005	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.49	13.38	13.29	13.21	13.15	13.09	13.03	12.98	12.94	12.90	12.66	12.53	12.40	12.27
5	0.01	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.96	9.89	9.82	9.77	9.72	9.68	9.64	9.61	9.58	9.55	9.38	9.29	9.20	9.11
5	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	6.52	6.49	6.46	6.43	6.40	6.38	6.36	6.34	6.33	6.23	6.18	6.12	6.07
5	0.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58	4.57	4.56	4.50	4.46	4.43	4.40
5	0.975	0.00	0.03	0.07	0.11	0.14	0.17	0.19	0.21	0.22	0.24	0.25	0.26	0.27	0.27	0.28	0.29	0.29	0.30	0.30	0.30	0.33	0.34	0.36	0.37
5	0.95	0.00	0.05	0.11	0.16	0.20	0.23	0.25	0.27	0.29	0.30	0.31	0.32	0.33	0.34	0.34	0.35	0.36	0.36	0.36	0.37	0.39	0.41	0.42	0.44
5	0.995	0.00	0.01	0.02	0.04	0.07	0.09	0.11	0.12	0.13	0.15	0.16	0.16	0.17	0.18	0.19	0.19	0.20	0.20	0.21	0.21	0.24	0.25	0.27	0.28
6	0.005	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.13	10.03	9.95	9.88	9.81	9.76	9.71	9.66	9.62	9.59	9.36	9.24	9.12	9.00
6	0.01	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.60	7.56	7.52	7.48	7.45	7.42	7.40	7.23	7.14	7.06	6.97
6	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41	5.37	5.33	5.30	5.27	5.24	5.22	5.20	5.18	5.17	5.07	5.01	4.96	4.90
6	0.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90	3.88	3.87	3.81	3.77	3.74	3.70
6	0.975	0.00	0.03	0.07	0.11	0.14	0.17	0.20	0.21	0.23	0.25	0.26	0.27	0.28	0.29	0.29	0.30	0.31	0.31	0.32	0.32	0.35	0.36	0.38	0.40
6	0.95	0.00	0.05	0.11	0.16	0.20	0.23	0.26	0.28	0.30	0.31	0.32	0.33	0.34	0.35	0.36	0.36	0.37	0.38	0.38	0.38	0.41	0.43	0.44	0.46
6	0.995	0.00	0.01	0.02	0.05	0.07	0.09	0.11	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20	0.20	0.21	0.21	0.22	0.22	0.25	0.27	0.29	0.30
7	0.005	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.27	8.18	8.10	8.03	7.97	7.91	7.87	7.83	7.79	7.75	7.53	7.42	7.31	7.19
7	0.01	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41	6.36	6.31	6.28	6.24	6.21	6.18	6.16	5.99	5.91	5.82	5.74
7	0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.71	4.67	4.63	4.60	4.57	4.54	4.52	4.50	4.48	4.47	4.36	4.31	4.25	4.20
7	0.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47	3.46	3.44	3.38	3.34	3.30	3.27
7	0.975	0.00	0.03	0.07	0.11	0.15	0.18	0.20	0.22	0.24	0.25	0.27	0.28	0.29	0.30	0.30	0.31	0.32	0.32	0.33	0.33	0.36	0.38	0.40	0.42
7	0.95	0.00	0.05	0.11	0.16	0.21	0.24	0.26	0.29	0.30	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.38	0.39	0.39	0.40	0.43	0.44	0.46	0.48
7	0.995	0.00	0.01	0.02	0.05	0.07	0.09	0.11	0.13	0.15	0.16	0.17	0.18	0.19	0.20	0.21	0.21	0.22	0.22	0.23	0.23	0.27	0.28	0.30	0.32
8	0.005	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.10	7.01	6.94	6.87	6.81	6.76	6.72	6.68	6.64	6.61	6.40	6.29	6.18	6.06
8	0.01	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61	5.56	5.52	5.48	5.44	5.41	5.38	5.36	5.20	5.12	5.03	4.95

TABLA F (Continuación)

		n_1																						
n_2	α	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	30	40	60	120
8	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.24	4.20	4.16	4.13	4.10	4.08	4.05	4.03	4.02	4.00	3.89	3.84	3.78	3.73
8	0.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17	3.16	3.15	3.08	3.04	3.01	2.97
8	0.975	0.00	0.03	0.07	0.11	0.15	0.18	0.20	0.23	0.24	0.26	0.27	0.28	0.30	0.30	0.31	0.32	0.33	0.33	0.34	0.34	0.38	0.40	0.41	0.43
8	0.95	0.00	0.05	0.11	0.17	0.21	0.24	0.27	0.29	0.31	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.39	0.40	0.40	0.41	0.44	0.46	0.48	0.50
8	0.995	0.00	0.01	0.02	0.05	0.07	0.09	0.12	0.13	0.15	0.16	0.18	0.19	0.20	0.21	0.21	0.22	0.23	0.23	0.24	0.24	0.28	0.30	0.32	0.34
9	0.01	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	5.05	5.01	4.96	4.92	4.89	4.86	4.83	4.81	4.65	4.57	4.48	4.40
9	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.91	3.87	3.83	3.80	3.77	3.74	3.72	3.70	3.68	3.67	3.56	3.51	3.45	3.39
9	0.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95	2.94	2.86	2.83	2.79	2.75
9	0.975	0.00	0.03	0.07	0.11	0.15	0.18	0.21	0.23	0.25	0.26	0.28	0.29	0.30	0.31	0.32	0.33	0.34	0.34	0.35	0.35	0.39	0.41	0.43	0.45
9	0.95	0.00	0.05	0.11	0.17	0.21	0.24	0.27	0.30	0.31	0.33	0.35	0.36	0.37	0.38	0.39	0.39	0.40	0.41	0.41	0.42	0.45	0.47	0.49	0.51
9	0.995	0.00	0.01	0.02	0.05	0.07	0.10	0.12	0.14	0.15	0.17	0.18	0.19	0.20	0.21	0.22	0.23	0.24	0.24	0.25	0.25	0.29	0.31	0.33	0.36
10	0.005	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.75	5.66	5.59	5.53	5.47	5.42	5.38	5.34	5.31	5.27	5.07	4.97	4.86	4.75
10	0.01	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60	4.56	4.52	4.49	4.46	4.43	4.41	4.25	4.17	4.08	4.00
10	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.66	3.62	3.58	3.55	3.52	3.50	3.47	3.45	3.44	3.42	3.31	3.26	3.20	3.14
10	0.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80	2.79	2.77	2.70	2.66	2.62	2.58
10	0.975	0.00	0.03	0.07	0.11	0.15	0.18	0.21	0.23	0.25	0.27	0.28	0.30	0.31	0.32	0.33	0.33	0.34	0.35	0.35	0.36	0.40	0.42	0.44	0.46
10	0.95	0.00	0.05	0.11	0.17	0.21	0.25	0.27	0.30	0.32	0.34	0.35	0.36	0.37	0.38	0.39	0.40	0.41	0.41	0.42	0.43	0.46	0.48	0.50	0.52
10	0.995	0.00	0.01	0.02	0.05	0.07	0.10	0.12	0.14	0.16	0.17	0.18	0.20	0.21	0.22	0.23	0.23	0.24	0.25	0.25	0.26	0.30	0.32	0.34	0.37
11	0.005	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.32	5.24	5.16	5.10	5.05	5.00	4.96	4.92	4.89	4.86	4.65	4.55	4.45	4.34
11	0.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.34	4.29	4.25	4.21	4.18	4.15	4.12	4.10	3.94	3.86	3.78	3.69
11	0.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.47	3.43	3.39	3.36	3.33	3.30	3.28	3.26	3.24	3.23	3.12	3.06	3.00	2.94
11	0.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67	2.66	2.65	2.57	2.53	2.49	2.45
11	0.975	0.00	0.03	0.07	0.11	0.15	0.18	0.21	0.24	0.26	0.27	0.29	0.30	0.31	0.32	0.33	0.34	0.35	0.36	0.36	0.37	0.41	0.43	0.45	0.48
11	0.95	0.00	0.05	0.11	0.17	0.21	0.25	0.28	0.30	0.32	0.34	0.35	0.37	0.38	0.39	0.40	0.41	0.41	0.42	0.43	0.43	0.47	0.49	0.51	0.53
11	0.995	0.00	0.01	0.02	0.05	0.07	0.10	0.12	0.14	0.16	0.17	0.19	0.20	0.21	0.22	0.23	0.24	0.25	0.25	0.26	0.27	0.31	0.33	0.36	0.38
12	0.005	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.99	4.91	4.84	4.77	4.72	4.67	4.63	4.59	4.56	4.53	4.33	4.23	4.12	4.01
12	0.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05	4.01	3.97	3.94	3.91	3.88	3.86	3.70	3.62	3.54	3.45
12	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.32	3.28	3.24	3.21	3.18	3.15	3.13	3.11	3.09	3.07	2.96	2.91	2.85	2.79
12	0.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57	2.56	2.54	2.47	2.43	2.38	2.34
12	0.975	0.00	0.03	0.07	0.11	0.15	0.19	0.21	0.24	0.26	0.28	0.29	0.31	0.32	0.33	0.34	0.35	0.35	0.36	0.37	0.37	0.41	0.44	0.46	0.49
12	0.95	0.00	0.05	0.11	0.17	0.21	0.25	0.28	0.30	0.33	0.34	0.36	0.37	0.38	0.39	0.40	0.41	0.42	0.43	0.43	0.44	0.48	0.50	0.52	0.55
12	0.995	0.00	0.01	0.02	0.05	0.07	0.10	0.12	0.14	0.16	0.18	0.19	0.20	0.22	0.23	0.24	0.24	0.25	0.26	0.27	0.27	0.31	0.34	0.36	0.39

TABLA F (Continuación)

		n1																							
n2	α	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	30	40	60	120
13	0.005	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.72	4.64	4.57	4.51	4.46	4.41	4.37	4.33	4.30	4.27	4.07	3.97	3.87	3.76
13	0.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.91	3.86	3.82	3.78	3.75	3.72	3.69	3.66	3.51	3.43	3.34	3.25
13	0.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.95	2.84	2.78	2.72	2.66
13	0.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48	2.47	2.46	2.38	2.34	2.30	2.25
13	0.975	0.00	0.03	0.07	0.11	0.15	0.19	0.22	0.24	0.26	0.28	0.29	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.37	0.38	0.42	0.44	0.47	0.50
13	0.95	0.00	0.05	0.11	0.17	0.21	0.25	0.28	0.31	0.33	0.35	0.36	0.38	0.39	0.40	0.41	0.42	0.42	0.43	0.44	0.44	0.48	0.51	0.53	0.55
13	0.995	0.00	0.01	0.02	0.05	0.08	0.10	0.12	0.14	0.16	0.18	0.19	0.21	0.22	0.23	0.24	0.25	0.26	0.26	0.27	0.28	0.32	0.35	0.37	0.40
14	0.005	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.51	4.43	4.36	4.30	4.25	4.20	4.16	4.12	4.09	4.06	3.86	3.76	3.66	3.55
14	0.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75	3.70	3.66	3.62	3.59	3.56	3.53	3.51	3.35	3.27	3.18	3.09
14	0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.09	3.05	3.01	2.98	2.95	2.92	2.90	2.88	2.86	2.84	2.73	2.67	2.61	2.55
14	0.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41	2.40	2.39	2.31	2.27	2.22	2.18
14	0.975	0.00	0.03	0.07	0.12	0.15	0.19	0.22	0.24	0.26	0.28	0.30	0.31	0.32	0.34	0.35	0.35	0.36	0.37	0.38	0.38	0.43	0.45	0.48	0.51
14	0.95	0.00	0.05	0.11	0.17	0.22	0.25	0.28	0.31	0.33	0.35	0.37	0.38	0.39	0.40	0.41	0.42	0.43	0.44	0.44	0.45	0.49	0.51	0.54	0.56
14	0.995	0.00	0.01	0.02	0.05	0.08	0.10	0.12	0.15	0.16	0.18	0.20	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.27	0.28	0.33	0.35	0.38	0.41
15	0.005	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.33	4.25	4.18	4.12	4.07	4.02	3.98	3.95	3.91	3.88	3.69	3.59	3.48	3.37
15	0.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56	3.52	3.49	3.45	3.42	3.40	3.37	3.21	3.13	3.05	2.96
15	0.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	3.01	2.96	2.92	2.89	2.86	2.84	2.81	2.79	2.77	2.76	2.64	2.59	2.52	2.46
15	0.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35	2.34	2.33	2.25	2.20	2.16	2.11
15	0.975	0.00	0.03	0.07	0.12	0.16	0.19	0.22	0.24	0.27	0.28	0.30	0.31	0.33	0.34	0.35	0.36	0.37	0.37	0.38	0.39	0.43	0.46	0.49	0.51
15	0.95	0.00	0.05	0.11	0.17	0.22	0.25	0.28	0.31	0.33	0.35	0.37	0.38	0.39	0.41	0.42	0.43	0.43	0.44	0.45	0.45	0.50	0.52	0.54	0.57
15	0.995	0.00	0.01	0.02	0.05	0.08	0.10	0.13	0.15	0.17	0.18	0.20	0.21	0.22	0.24	0.25	0.26	0.26	0.27	0.28	0.29	0.33	0.36	0.39	0.42
16	0.005	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.18	4.10	4.03	3.97	3.92	3.87	3.83	3.80	3.76	3.73	3.54	3.44	3.33	3.22
16	0.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.50	3.45	3.41	3.37	3.34	3.31	3.28	3.26	3.10	3.02	2.93	2.84
16	0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.93	2.89	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.68	2.57	2.51	2.45	2.38
16	0.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30	2.29	2.28	2.19	2.15	2.11	2.06
16	0.975	0.00	0.03	0.07	0.12	0.16	0.19	0.22	0.25	0.27	0.29	0.30	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.39	0.44	0.46	0.49	0.52
16	0.95	0.00	0.05	0.12	0.17	0.22	0.25	0.29	0.31	0.33	0.35	0.37	0.38	0.40	0.41	0.42	0.43	0.44	0.44	0.45	0.46	0.50	0.53	0.55	0.58
16	0.995	0.00	0.01	0.02	0.05	0.08	0.10	0.13	0.15	0.17	0.18	0.20	0.21	0.23	0.24	0.25	0.26	0.27	0.27	0.28	0.29	0.34	0.37	0.40	0.43
17	0.005	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	4.05	3.97	3.90	3.84	3.79	3.75	3.71	3.67	3.64	3.61	3.41	3.31	3.21	3.10
17	0.01	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.40	3.35	3.31	3.27	3.24	3.21	3.19	3.16	3.00	2.92	2.83	2.75
17	0.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.87	2.82	2.79	2.75	2.72	2.70	2.67	2.65	2.63	2.62	2.50	2.44	2.38	2.32

TABLA F (Continuación)

		n_1																					
n_2	α	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	30	40	60	120
17	0.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26	2.24	2.23	2.15	2.10	2.06	2.01
17	0.975	0.00	0.03	0.07	0.12	0.16	0.19	0.22	0.25	0.27	0.29	0.30	0.32	0.33	0.34	0.36	0.37	0.37	0.38	0.39	0.40	0.44	0.47	0.50	0.53
17	0.95	0.00	0.05	0.12	0.17	0.22	0.26	0.29	0.31	0.34	0.36	0.37	0.39	0.40	0.41	0.42	0.43	0.44	0.45	0.46	0.46	0.51	0.53	0.56	0.59
17	0.995	0.00	0.01	0.02	0.05	0.08	0.10	0.13	0.15	0.17	0.19	0.20	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.29	0.34	0.37	0.40	0.44
18	0.005	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.94	3.86	3.79	3.73	3.68	3.64	3.60	3.56	3.53	3.50	3.30	3.20	3.10	2.99
18	0.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.32	3.27	3.23	3.19	3.16	3.13	3.10	3.08	2.92	2.84	2.75	2.66
18	0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.81	2.77	2.73	2.70	2.67	2.64	2.62	2.60	2.58	2.56	2.44	2.38	2.32	2.26
18	0.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22	2.20	2.19	2.11	2.06	2.02	1.97
18	0.975	0.00	0.03	0.07	0.12	0.16	0.19	0.22	0.25	0.27	0.29	0.31	0.32	0.34	0.35	0.36	0.37	0.38	0.39	0.39	0.40	0.45	0.47	0.50	0.54
18	0.95	0.00	0.05	0.12	0.17	0.22	0.26	0.29	0.32	0.34	0.36	0.37	0.39	0.40	0.41	0.42	0.43	0.44	0.45	0.46	0.46	0.51	0.54	0.56	0.59
18	0.995	0.00	0.01	0.02	0.05	0.08	0.10	0.13	0.15	0.17	0.19	0.20	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30	0.35	0.38	0.41	0.44
19	0.005	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.84	3.76	3.70	3.64	3.59	3.54	3.50	3.46	3.43	3.40	3.21	3.11	3.00	2.89
19	0.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24	3.19	3.15	3.12	3.08	3.05	3.03	3.00	2.84	2.76	2.67	2.58
19	0.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.76	2.72	2.68	2.65	2.62	2.59	2.57	2.55	2.53	2.51	2.39	2.33	2.27	2.20
19	0.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18	2.17	2.16	2.07	2.03	1.98	1.93
19	0.975	0.00	0.03	0.07	0.12	0.16	0.19	0.22	0.25	0.27	0.29	0.31	0.32	0.34	0.35	0.36	0.37	0.38	0.39	0.40	0.40	0.45	0.48	0.51	0.54
19	0.95	0.00	0.05	0.12	0.17	0.22	0.26	0.29	0.32	0.34	0.36	0.38	0.39	0.40	0.42	0.43	0.44	0.45	0.45	0.46	0.47	0.51	0.54	0.57	0.60
19	0.995	0.00	0.01	0.02	0.05	0.08	0.10	0.13	0.15	0.17	0.19	0.20	0.22	0.23	0.24	0.26	0.27	0.27	0.28	0.29	0.30	0.35	0.38	0.41	0.45
20	0.005	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.76	3.68	3.61	3.55	3.50	3.46	3.42	3.38	3.35	3.32	3.12	3.02	2.92	2.81
20	0.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09	3.05	3.02	2.99	2.96	2.94	2.78	2.69	2.61	2.52
20	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.72	2.68	2.64	2.60	2.57	2.55	2.52	2.50	2.48	2.46	2.35	2.29	2.22	2.16
20	0.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15	2.14	2.12	2.04	1.99	1.95	1.90
20	0.975	0.00	0.03	0.07	0.12	0.16	0.19	0.22	0.25	0.27	0.29	0.31	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40	0.41	0.46	0.48	0.51	0.55
20	0.95	0.00	0.05	0.12	0.17	0.22	0.26	0.29	0.32	0.34	0.36	0.38	0.39	0.41	0.42	0.43	0.44	0.45	0.46	0.46	0.47	0.52	0.54	0.57	0.60
20	0.995	0.00	0.01	0.02	0.05	0.08	0.10	0.13	0.15	0.17	0.19	0.21	0.22	0.23	0.25	0.26	0.27	0.28	0.29	0.29	0.30	0.35	0.38	0.42	0.46
30	0.005	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.25	3.18	3.11	3.06	3.01	2.96	2.92	2.89	2.85	2.82	2.63	2.52	2.42	2.30
30	0.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74	2.70	2.66	2.63	2.60	2.57	2.55	2.39	2.30	2.21	2.11
30	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.46	2.41	2.37	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.07	2.01	1.94	1.87

TABLA F (Continuación)

		n_1																							
n_2	**α**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**30**	**40**	**60**	**120**
30	**0.05**	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96	1.95	1.93	1.84	1.79	1.74	1.68
30	**0.975**	0.00	0.03	0.07	0.12	0.16	0.20	0.23	0.26	0.28	0.30	0.32	0.34	0.35	0.37	0.38	0.39	0.40	0.41	0.42	0.43	0.48	0.51	0.55	0.59
30	**0.95**	0.00	0.05	0.12	0.17	0.22	0.26	0.30	0.32	0.35	0.37	0.39	0.41	0.42	0.43	0.45	0.46	0.47	0.47	0.48	0.49	0.54	0.57	0.61	0.64
30	**0.995**	0.00	0.01	0.02	0.05	0.08	0.11	0.13	0.16	0.18	0.20	0.21	0.23	0.25	0.26	0.27	0.28	0.29	0.30	0.31	0.32	0.38	0.42	0.46	0.50
40	**0.005**	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	3.03	2.95	2.89	2.83	2.78	2.74	2.70	2.66	2.63	2.60	2.40	2.30	2.18	2.06
40	**0.01**	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52	2.48	2.45	2.42	2.39	2.37	2.20	2.11	2.02	1.92
40	**0.025**	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.13	2.11	2.09	2.07	1.94	1.88	1.80	1.72
40	**0.05**	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92	1.90	1.89	1.87	1.85	1.84	1.74	1.69	1.64	1.58
40	**0.975**	0.00	0.03	0.07	0.12	0.16	0.20	0.23	0.26	0.29	0.31	0.33	0.34	0.36	0.37	0.39	0.40	0.41	0.42	0.43	0.44	0.50	0.53	0.57	0.62
40	**0.95**	0.00	0.05	0.12	0.17	0.22	0.26	0.30	0.33	0.35	0.38	0.40	0.41	0.43	0.44	0.45	0.46	0.48	0.48	0.49	0.50	0.56	0.59	0.63	0.67
40	**0.995**	0.00	0.01	0.02	0.05	0.08	0.11	0.13	0.16	0.18	0.20	0.22	0.24	0.25	0.27	0.28	0.29	0.30	0.31	0.32	0.33	0.40	0.44	0.48	0.53
60	**0.005**	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.82	2.74	2.68	2.62	2.57	2.53	2.49	2.45	2.42	2.39	2.19	2.08	1.96	1.83
60	**0.01**	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.03	1.94	1.84	1.73
60	**0.025**	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.22	2.17	2.13	2.09	2.06	2.03	2.01	1.98	1.96	1.94	1.82	1.74	1.67	1.58
60	**0.05**	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86	1.84	1.82	1.80	1.78	1.76	1.75	1.65	1.59	1.53	1.47
60	**0.975**	0.00	0.03	0.07	0.12	0.16	0.20	0.24	0.26	0.29	0.31	0.33	0.35	0.37	0.38	0.40	0.41	0.42	0.43	0.44	0.45	0.52	0.55	0.60	0.65
60	**0.95**	0.00	0.05	0.12	0.18	0.23	0.27	0.30	0.33	0.36	0.38	0.40	0.42	0.44	0.45	0.46	0.47	0.49	0.50	0.51	0.51	0.57	0.61	0.65	0.70
60	**0.995**	0.00	0.01	0.02	0.05	0.08	0.11	0.14	0.16	0.18	0.21	0.22	0.24	0.26	0.27	0.29	0.30	0.31	0.32	0.33	0.34	0.41	0.46	0.51	0.57
120	**0.005**	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.62	2.54	2.48	2.42	2.37	2.33	2.29	2.25	2.22	2.19	1.98	1.87	1.75	1.61
120	**0.01**	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34	2.28	2.23	2.19	2.15	2.12	2.09	2.06	2.03	1.86	1.76	1.66	1.53
120	**0.025**	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.10	2.05	2.01	1.98	1.94	1.92	1.89	1.87	1.84	1.82	1.69	1.61	1.53	1.43
120	**0.05**	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.80	1.78	1.75	1.73	1.71	1.69	1.67	1.66	1.55	1.50	1.43	1.35
120	**0.975**	0.00	0.03	0.07	0.12	0.16	0.20	0.24	0.27	0.29	0.32	0.34	0.36	0.38	0.39	0.41	0.42	0.43	0.44	0.45	0.46	0.54	0.58	0.63	0.70
120	**0.95**	0.00	0.05	0.12	0.18	0.23	0.27	0.31	0.34	0.36	0.39	0.41	0.43	0.44	0.46	0.47	0.49	0.50	0.51	0.52	0.53	0.59	0.63	0.68	0.74
120	**0.995**	0.00	0.01	0.02	0.05	0.08	0.11	0.14	0.16	0.19	0.21	0.23	0.25	0.27	0.28	0.30	0.31	0.32	0.33	0.35	0.36	0.43	0.48	0.55	0.62

BIBLIOGRAFÍA

- Aliaga, M. y Gunderson B. (1998). *"Interactive Statistics".* Edition Preliminary. Prentice Hall. Inc.

- Agresti, Alan (1990). *"Categorical Data Analysis"*. John Wiley & Sons, Inc.

- Agresti, Alan (1996). *"An Introduction to Categorical Data Analysis"*. John Wiley & Sons, Inc.

- Ato García, M. y Lopez García, J.J (1996). *"Análisis Estadístico para Datos Categóricos"*. Síntesis S.A. Madrid.

- Cook, R.D y Weisberg.S. (1986). *"Residuals and Influence in Regression"* Chapman and Hall.

- Dixon, W. y Massey, F.J. (1970). *"Introducción al Análisis Estadístico".* McGraw-Hill.Inc. México.

- Freund, J. y Manning Smith, R. (1989). *"Estadística"*. Cuarta Edición. Prentice Hall Hispanoamericana S.A. México.

- Freund, J. y Walpole, R. (1990). *"Estadística Matemática con Aplicaciones".* Cuarta Edición. Prentice Hall Hispanoamericana S.A. México.

- Guzman M. ; Colera J. y Salvador A. (1987). *"Matemáticas: Bachillerato I". Ed. Anaya S.A. Madrid.*

- Harnett, D. y Murphy, J. (1987). *"Introducción al Análisis Estadístico".* Segunda Edición. Addison-Wesley Iberoamericana.S. A.

- Johnson N. L. y Kotz S. (1970). *"Continuous Univariate Distributions - 2"*. Houghton Mifflin Company. Boston.

- Lehmann E.L. (1983) . *"Theory of Point Estimation"* . John Wiley & Sons, Inc.

- Martinez, Ricardo (1995). *" Diseños Aplicados en Frutales y en Biotecnología"*
Simposio Internacional de Estadística. Departamento de Matemáticas y Estadística. Universidad Nacional de Colombia.

- Mead, R. , Curnow R. N. y Hasted, A.M. (1993). *"Statistical Methods in Agriculture and Experimental Biology"* . Chapman & Hall.

- Mendenhall, W. ; Wackerly, D. y Scheaffer, R. (1994). *"Estadística Matemática con Aplicaciones"* . Segunda Edición. Grupo Editorial Iberoamerica.México.

- Meyer, Paul. (1992). *"Probabilidad y Aplicaciones Estadísticas"* Edición Revisada. Addison-Wesley Iberoamericana.México.

- Mischan, M. y Zambello de Pinho, S. (1996). "Experimentação Agronômica - Dados Não-Balanceados" FUNDIBIO. Botucatu. San Pablo. Brasil.

- Montgomery, D. y Peck, E. (1982). *“Introduction to Linear Regression Analysis”*. John Wiley & Sons, Inc.

- Peña Sanchez de Rivero D. (1988). *“Estadística Modelos y Métodos”* Segunda Edición. Alianza Editorial Textos Madrid.

- Snedecor, G. W. y Cochran, W.G. (1978). *“Métodos Estadísticos”*. Quinta Edición. Compañía Editorial Continental S.A. México.

- Software Microsoft *Excel'97*

- Sokal, R. y Rohlf, J. (1980) *“Introducción a la Bioestadística”.* Editorial Reverté. S.A. Barcelona España.

- Spiegel, Murray R. (1991) *“Estadística”* Segunda Edición. McGRAW-HILL. Madrid.

- Steel, R.G.D. y Torrie J. (1985). *“Bioestadística: Principios y Procedimientos”* . Segunda Edición. McGraw-Hill.Inc. Latinoamericana S.A. Bogotá Colombia.

- Ya Lung Chou (1977). *”Análisis Estadístico”.* Interamericana. México.



Introducción a la estadística para las ciencias de la vida

Elsa Moschetti, Susana Ferrero
Gabriela Palacio y Marcelo Ruiz

Las ciencias biológicas y la estadística son dos lenguajes que históricamente se han desarrollado en estrecha relación. El presente texto, atendiendo a este vínculo –a partir de problematizaciones derivadas del campo biológico- intenta introducir al estudiante a la metodología y teoría estadística.
Se ha tratado de respetar los contextos reales de surgimiento de los problemas de diferentes disciplinas (veterinaria, agronomía, biología, etc.) y al mismo tiempo mantener la unidad de la técnica y su funcionamiento. Al respecto, se pretende evitar tanto la "receta instrumental" como el formalismo poco útil.
Dado que el texto presenta un desarrollo autocontenido de los conceptos y relaciones, para su comprensión solo son necesarios conocimientos elementales de matemática.

Autores

Los autores son docentes del Departamento de Matemática de la facultad de Ciencias Exactas, Físico-Químicas y Naturales de la Universidad Nacional de Río Cuarto y han desempeñado sus tareas docentes en el área de estadística de dicho departamento. El libro es el resultado de un proceso que ha involucrado tanto a la actividad docente, en diferentes asignaturas relacionadas con la estadística, como a la investigación y asesoramiento estadístico en el mismo campo disciplinar.

Colección Académico-Científica
U*a+U





